CHARACTERISATION OF LIPOPROTEINS IN *STAPHYLOCOCCUS AUREUS*

ASHRAF ELGALLALI

Centre for Parasitology and Disease Research,

School of Environment and Life Sciences,

University of Salford,

Salford, UK

Submitted In Partial Fulfilment of the Requirement of the Degree of Doctor of Philosophy, 2016

Table of Contents

List of tables

List of figures

Acknowledgements

I am extremely indebted to Allah for the good health and wellbeing that were necessary to complete this thesis.

I would sincerely like to express my special appreciation and thanks to my supervisor, Prof. Howard Foster for his incredible endless support and his continuous willingness and assistance to finish my PhD, for his patience, motivation and enormous knowledge, without his guidance and encouragement this thesis would not be completed.

Very special thanks to my family: my parents, brother and sisters for supporting me, without their motivation and encouragement I would not complete this thesis.

I would especially like to thank my friends, I am extremely grateful and indebted to them for sharing expertise, and valuable guidance and encouragement extended to me.

I would also thank all of my friends and colleagues in the lab for maintaining a good and nice work atmosphere.

Declaration

No portion of this work referred to in this report has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

List of abbreviations

aa …… amino acid

Agr ……. Accessory gene regulator

CA-MRSA……..…….… Community-acquired methicillin-resistant *S. aureus*

CHIP…….…. Chemotaxis inhibitory protein

CPM ……... Counts Per million Mapped reads

DE …….. Differentially Expressed

DOLOP …….. Database of Bacterial Lipoproteins

DNase ……. Deoxyribonuclease

DTT …… Dithiotreitol

EAp……..… Extracellular Adherence protein

E-value ……Expect value

FC ……Fold Change

FDR …… False Discovery Rate

FPKM ……. Fragments per Kilobase Transcript per Million mapped reads

Fur ……. Ferric uptake regulator

GFP ……..... Green fluorescent protein

HA-MRSA…….…. Hospital-acquired methicillin-resistant *Staphylococcus aureus*

IM lipoproteins ……..……. Inner-Membrane lipoproteins

IAA ……. Iodoacetamide

ICU ….…… Intensive Care Units

ISA ……..……Iso-Sensitest agar

Isd……...…… Iron-regulated surface determinant

pro-IL-1β …….pro-Interleukin-1β

LC ……..……. Liquid Chromatography

LGT..…….…….. Lateral Gene Transfer

Lpp ……...….. Lipoproteins

LPXTG motif …….. (Leu-Pro-any-Thr-Gly)

Lsp ……... Lipoprotein signal peptidase

MALP-2 ……... Macrophage activating lipopeptide-2

MRSA ……..…... Methicillin-resistant *Staphylococcus aureus*

MSCRAMM……..... Microbial surface components recognizing adhesive matrix molecules

NJ ……...…….. Neighbor joining tree

NOD……..…… Nucleotide-binding oligomerization domain

OM lipoproteins ……..... Outer Membrane lipoproteins

ORFs ……..…..… Open reading frames

OTU ……..……... Operational taxonomic unit

PAMP……..…….. Pathogen-associated molecular pattern

PBP…… Penicillin-binding protein

PMNs cells….……..….. Poly morphonuclear leukocyte

PRR…… Pattern recognition receptor

PSM……..…. Phenol Soluble Modulin

PV ……..…... P-value

PVL ……..……Panton-Valentin leucocidin

RNase ……..ribonuclease

SAB ……...… *Staphylococcus aureus* bacteraemia

SBPs ……. Substrate Binding Proteins

SCC*mec* ……... Staphylococcal chromosomal cassette *mec*

SDS ….…….……...….. Sodium Dodecyl Sulfate

SAR ……...…Staphylococcal Accessory Regulator

SNPs ……..... Single-nucleotide polymorphism

SP……...Signal peptide

SOD……...…… Superoxide dismutase

srtA …….. surface protein sorting A

Tat ……..….. Twin arginine translocation pathway

TCR…….…..…T cell receptor

TMED …..…….…….. Tetramethylethylenediamine

TLR…..…..……..…Toll-like receptor

TSB…..……...…….….. Tryptic Soy Broth

TST ……... Toxic Shock Toxin-1

VRSA .....….. vancomycin resistant *Staphylococcus aureus*

VISA …….…… vancomycin intermediate *Staphylococcus aureus*

WT……..… Wild type

2-DE …..……..Two-dimensional gel electrophoresis

Abstract

The Gram-positive bacterium *Staphylococcus aureus* is an extremely successful opportunistic bacterium capable of causing a wide range of hospital-acquired and community-acquired infections, and is becoming increasingly virulent and resistant to antibiotics. In order to investigate this pathogen, various methods have been used to analyse the pathogenic behaviour including genomics, transcriptomics and proteomics. *S. aureus* expresses approximately 55-70 lipoproteins with only about half with known functions. Little is known about the biochemical functions of many individual lipoproteins and their proteomics has not been investigated in detail. Lipoproteins have a broad ranging functionality and perform various roles in bacterial activity and attract a particular interest to investigate their virulence and survival influences in the course of host infection. The initial part of this study was to find out whether the lipoproteins of *S. aureus* have similar genetic characteristics among all strains. PCR and Quantitative Real-Time PCR experiments were performed to analyse the genetic and the expression levels for some lipoprotein genes. The majority of PCR results showed high similarity in lipoprotein genetic structure among the examined strains. Phylogenetic trees from concatenated lipoprotein genes alignment were generated to represent the lipoprotein genes distribution of *S. aureus* strains. To identify and characterise proteomic of *S. aureus* lipoproteins a comprehensive quantitative proteome profiling of *S. aureus* lipoproteins using gel-free /in-solution trypsin digestion system followed by LC-MS/MS quantification identified 38 lipoproteins that represent two-thirds of the *S. aureus* MRSA252 lipoprotein. In addition, *S. aureus*-mediated infections with live *C. elegans* were performed on solid assays to investigate the host-pathogen relationships. *S. aureus* MRSA252 exhibited a high level of nematocidal activity with average time for half of the worms to die of ~ 2 d and infected *C. elegans* showed visible signs of illness. To evaluate lipoprotein transcripts expression level and microbe/host-specific pathogenic factors RNA of both *S. aureus* and *C. elegans* were characterised after isolation from the infected *C. elegans* and subjected to RNA Sequencing, the large-scale data has provided useful information on pathogen and host activities during infection. RNA sequencing analysis showed different types of regulations and interactions of lipoprotein transcripts during host exposures to indicate 3 transcripts significantly were up-regulated and 11 down-regulated. RNA sequencing analysis showed that 62 lipoprotein transcripts were expressed during *C. elegans* infection model. Proteomic analysis using the application of gel-free proteomic technique identified 38

lipoproteins that were expressed in the non-infection condition representing approx. two-thirds of the *S. aureus* MRSA252 lipoproteins. The results suggest that some lipoproteins were involved in pathogenesis of *C. elegans* but their function were not clear. More research is needed for explore the roles of lipoproteins in pathogenesis and the interactions of *S. aureus* with the host immune responses.

Chapter one

Introduction and the aims of this thesis

1. Introduction

1.1 *Staphylococcus aureus*

1.1.1 General introduction

Staphylococcus aureus has been recognized as one of the most common pathogenic Gram-positive bacteria around the world with a serious prevalence of hospital- and community-associated (CA) infections. It causes a range of infectious diseases such as sepsis, endocarditis and pneumonia. In addition, it can cause uncomplicated skin infections impetigo, also soft tissue infections, with or without abscess formation (Noguchi *et al.*, 2006). Furthermore, *S. aureus* can cause toxin-mediated diseases such as toxic shock syndrome, scalded skin syndrome and food poisoning (Dinges *et al.*, 2000). *S. aureus* are pathogenic bacteria that have become a great public health concern in recent years due to their ability to produce an array of virulence factors and drug-resistant variants. Before the availability of antibiotics, *S. aureus* infections were most often non-fatal but after the discovery and use of penicillin the prognosis of staphylococcal infections significantly improved. The first reported isolation of methicillin-resistant staphylococcus aureus in 1961 in the United Kingdom was shortly after methicillin introduced into clinical use (Barber, 1961). Since that time infections caused by methicillin resistant *S. aureus* (MRSA) have become causes of predominantly hospital-acquired and community-acquired infections with emergence of various antimicrobial resistances over the past years and causing high morbidity and mortality rates. Both epidemic potential strains EMRSA-15 and EMRSA-16 emerged in the United Kingdom in the early 1990s and causing life threatening infections (Johnson *et al.*, 2001). The ability of *S. aureus* to evolve new virulent strains and acquire resistance to a number of antibiotics including methicillin (methicillin-resistant *Staphylococcus aureus*; MRSA) has increased dramatically. However, these strains started to spread rapidly in the late 1980s to the mid-1990s, and have been known to be endemic in some hospitals and become more difficult to eradicate (Lindsay and Holden, 2004). MRSA prevalence in United Kingdom: <1% of patients living at home, 22% of care home residents and 40% of positive *S. aureus* blood cultures, 82% of those with MRSA infection are ≥ 60 years (Health Protection Agency, 2009). The surveillance on the cases of methicillin-resistant *S. aureus* blood stream infections (BSIs) recorded in England between 2003 and 2013, the numbers of reported MRSA BSIs in 2003-2004 was 7700 cases and such infections showed a notable decline after the

introduction of legislation and introduction of healthcare-associated improvement programmes in 2006 was significantly associated with the reduction of MRSA BSIs (Duerden *et al.*, 2015). The MRSA BSI rates per 100, 000 population have slightly decreased from 1.8 to 1.7 to 1.4 and to 1.5 in 2012, 2013, 2014 and 2015, respectively (Health Protection Agency, 2015). *S. aureus* related infections include skin and soft tissue infections, pneumonia, infective endocarditis and bloodstream infections in United States are over half a million cases and more than 10,000 deaths per year (Klevens *et al.*, 2007, Otto, 2010). Methicillin resistance in *S. aureus* is due to the acquisition of a mobile genetic element, their size ranging from 21/67 kb in size, termed the staphylococcal cassette chromosome *mec* (SCC*mec*) (Katayama *et al.*, 2000).

Glycopeptide antibiotics including vancomycin prevent maturation of bacterial cell wall by binding to the terminal D-alanyl-D-alanine residues of peptidoglycan precursors, consequently blocking the bacterial enzymes involved in the late steps of peptidoglycan synthesis (Anderson *et al.*, 1965). Infections with MRSA strains are usually treated with vancomycin, but the first vancomycin intermediate-level resistant isolates (VISA) were discovered in 1997 (Hiramatsu *et al.*, 1997). There are early reports of high-level vancomycin resistant isolates (VRSA) that have acquired the vanA resistance gene from vancomycin resistant enterococci (Control and Prevention, 1997, Hierholzer *et al.*, 1995). Reports revealed that 40%-60% of all hospital *S. aureus* infections in Europe, United States and Japan were resistant to methicillin (Lindsay and Holden, 2006). Resistant MRSA strains express a special enzyme known as penicillin-binding protein 2a (PBP2a) which is not inactivated by β-lactams and keeps generating the cross-linked peptidoglycan in their presence (Ubukata *et al.*, 1985). Platensimycin antibiotic has been shown to be effective against some strains of MRSA and VRSA (Wang *et al.*, 2007). There are some limited effective antibiotics that cure MRSA infections and some are to be developed in the near future (Dryden, 2014).

1.1.2 *S. aureus* microbiology and taxonomy

S. aureus is a member of the family *Micrococcaceae* and on microscopic examination the organisms appear as Gram-positive spherical bacteria (cocci) non-motile and non-spore forming with a diameter of 0.5–1.5 μm, appearing in clusters on the Gram's stain (Wesley and Bannerman, 1994). It is distinguished from other staphylococcal species with the gold pigmentation of their colonies when grown on solid media (β-haemolytic colonies), and

positive results of coagulase (Willis *et al.*, 1966), mannitol-fermentation and deoxyribonuclease tests (Raymond and Traub, 1970). It is catalase positive (unlike streptococci), a facultative anaerobic organism. *S. aureus* is extremely heat sensitive and is inactivated at a temperature > 46ºC while their enterotoxins are heat-stable proteins which survive heat treatment and low pH conditions (Evenson *et al.*, 1988). Both MRSA and methicillin-susceptible *S. aureus* have been found to survive for many weeks in a completely dry environment (Beard-Pegler *et al.*, 1988).

The *S. aureus* cell wall is a strong protective coat and relatively amorphous form, about 20-50nm in thickness (Shockman and Barren, 1983), composed from repeating disaccharide N-acetylmuramic acid-(β1–4)-N-acetylglucosamine, cross-linking is via a pentglycine interpeptide bridge (Ghuysen and Strominger, 1963). The staphylococcal cell wall also contains secondary polymers (proteins, carbohydrates and teichoic acids) that are immobilised in the peptidoglycan scaffold (Navarre and Schneewind, 1999). Peptidoglycan is the major component of the cell wall and comprises up to 50% of the cell wall mass (Waldvogel, 1995). However, there is another essential component group of phosphate containing polymers known as teichoic acids. *S. aureus* has two types of teichoic acids, cell wall teichoic acid and cell membrane associated lipoteichoic acid, both of them make up about 40% of cell wall mass (Knox and Wicken, 1973), which give a negative charge to the bacterial cell surface and take part in the acquisition and localisation of metal ions especially in divalent cations and the activities of autolytic enzymes (Wilkinson, 1997). Over 90% of *S. aureus* clinical isolated strains have been shown to possess capsular polysaccharides; production of this capsule is reported to decrease phagocytosis *in vitro* and to increase *S. aureus* virulence in a mouse bacteraemia model (Thakker *et al.*, 1998). *S. aureus* is typically an extracellular pathogen able to survival and persist in different host tissues (Brouillette *et al.*, 2003), it is also capable of internalising within non-professional phagocytic cells such as epithelial and endothelial cells for a short time, this strategy helps the pathogen to evade host defence mechanisms and the action of antimicrobial agents that mainly act in the extracellular space (Sinha *et al.*, 1999).

1.1.3 Transmission of *S. aureus*

About thirty percent of the human population are natural reservoirs for *S. aureus* by carrying the organism in their nose (in the anterior nares) (Wertheim *et al.*, 2005), 58% on their throat (Nilsson and Ripa, 2006) and hand carriage was found in 24.9% (David *et al.*,

2014), with asymptomatic colonisation being more frequent than infection. Throat colonization was more common than nasal colonization (Marshall and Spelman, 2007). Between 25% -50% of healthy persons may be persistently or transiently colonised, however, colonisation rate is higher amongst individuals who are immuno-compromised, for example diabetic patients, HIV-infected patients, patients who require haemodialysis and people with skin diseases (Sydnor and Perl, 2011). *S. aureus* also colonises and infects many mammals including domestic animals such as dogs and cats, some birds e.g. chickens and turkeys, and farm animals such as cows, pigs and goats (Baptiste *et al.*, 2005). Infection with *S. aureus* is a common cause of dairy cow mastitis with significant economic impact; moreover these animals may act as good reservoirs for human colonisation (McCarthy and Lindsay, 2010).

The majority of *S. aureus* patients with symptomatic infections are invaded with their own colonising strains (Kluytmans *et al.*, 1997). Infection might be also acquired from another person or from environmental exposure. *S. aureus* transmission in the community sector is associated with skin-to-skin contact and contaminated household environment (Desai *et al.*, 2011), bacteria can survive for long periods on household fomites for up to 2 months (Baggett *et al.*, 2004). Reports on MRSA environmental contamination have shown that 5%-8% among households without CA-MRSA disease however in households with a CA-MRSA-infected person or a healthcare labourer were 26%-32% (Scott *et al.*, 2008, Uhlemann *et al.*, 2011). Airborne spread and transmission through contacts with contaminated materials may also be involved (Cooper *et al.*, 2004).

Hospital transmission (patient to patient) is the most likely outcome from transient colonisation of the hands of hospital workers, who then transfer strains between patients. Spread of bacteria in aerosols from the respiratory or nasal discharges from heavily colonised individuals has been reported (David and Daum, 2010). *S. aureus* host switching is an important mechanism in the evolution of pathogenicity, for instance, transmission from bovine to human and from human to poultry host switches have been observed (Lowder *et al.*, 2009, Sakwinska *et al.*, 2011). The considered human commensal pathogen *S. aureus* has been isolated from numerous vertebrate species (Kloos, 1980). The main reservoir of *S. aureus* in the home is colonized or infected individuals and domestic animals (Bloomfield *et al.*, 2007).

Hygiene and medical interventions in United Kingdom have led to a significant reduction in incidence of health care-associated methicillin resistant *S. aureus* MRSA related infections (Mandatory Surveillance MRSA). *S. aureus* MRSA shows attributes that are not always found in most clinically isolated bacteria, it has the ability to express a range of virulence factors and therefore is always considered medically relevant when encountered in clinical specimens. *S. aureus* is able to develop and expand resistance to a wide range of antimicrobial treatments and it is a major pathogen in both hospital and community infections (Styers *et al.*, 2006).

1.1.4 Epidemiology of *S. aureus* infections

Development of new molecular typing approaches has allowed the study of population structure and epidemiology of bacterial pathogens, but these approaches are variable in practical for large population samples. The evolution of methicillin-resistant *Staphylococcus aureus* (MRSA) has changed the clinical and molecular epidemiology of *S. aureus* infections in the past three decades, these observations have showed the importance of epidemiological variations of *S. aureus* infections and the molecular characterisation of pathogenic resistant strains. Bacterial adaptability and the significant speed of bacterial evolution together with environmental challenges give this bacterium an ability to create genetic variation for their survival. Little information is available about the epidemiology of *S. aureus* in the non-western parts of the world; due to the increasing number of people traveling worldwide this epidemiology has changed.

In general, numbers of both community-acquired and hospital-acquired staphylococcal infection has increased with increasing levels of antibiotic resistance and the emergence of epidemic strains (Chatterjee and Otto, 2013). *S. aureus* pathogenicity indicates that virulence is a multi-factorial process as single gene inactivation experiments could not prevent *S. aureus* pathogenic ability (Fedtke *et al.*, 2004). The national data on hospitalizations and antibiotic resistance in *S. aureus* MRSA infections in the United States in 2010, the records show 4,476 MRSA colonized/infected patients in 67,412 inpatients with prevalence rate is 66.4 per 1,000 inpatients, 25.3 infections and 41.1 colonisations per 1,000 inpatients (Jarvis *et al.*, 2012). Attempts to reduce the prevalence of MRSA in hospital have achieved some success, in the period between 2001 and 2007, MRSA central line-associated blood stream infections within ICU cases in US were decreased by almost 50% (Burton *et al.*, 2009). Investigated study in hospitalized patients

between 2005 and 2008 confirmed a 34 % decreased of MRSA-related bloodstream infections incidence (Kallen *et al.*, 2010). *S. aureus* bacteraemia (SAB) infection overall rates may have stabilized over the past two decades, with incidence of SAB vary from 10 to 30 cases per 100,000 person per a year in industrialized community (Laupland *et al.*, 2013). A little is known about the incidence of SAB in the nonindustrialized, in contrast, incidence of community-acquired SAB between 2004 and 2010 in northeast Thailand was approx. 3 per 100,000 people-years (Kanoksil *et al.*, 2013). The incidence of MRSA cases among 3,662 *S. aureus* isolated in four hospitals in Siberian Russia between 2007 and 2011, prevalence of HA-MRSA was 22% while the CA-MRSA was 2.9% (Khokhlova *et al.*, 2015).

From a clinical point of view, new approaches to reduce the *S. aureus* infections include minimizing the duration of hospital stay, improved surveillance systems, use of antibiotic to eliminate nasal carriage before elective surgery and more strict hand hygiene regulations (Skov *et al.*, 2012). English National Point Prevalence Survey on Healthcare-associated Infections and Antimicrobial Use indicated a sharp decrease of healthcare-associated MRSA bacteraemia from 1.8% of cases with MRSA bacteraemia in 2006 to 0.1% in 2011 (Mandatory Surveillance MRSA). Over time, CA-MRSA epidemic became more complicated in various geographic locations and urgently need to improve the understanding of *S. aureus* epidemiology. Schaumburg *et al*. investigated the epidemiology of *S. aureus* isolates in the African continent in both rural communities and developed urban populations were found to contain the Panton-Valentine leukocidin gene that encodes the potent leukotoxin in 17% to 74% of isolates, the majority of these PVL-positive strains were methicillin-susceptible, suggested that PVL epidemiology in Africa was different from that in Europe where the prevalence of *S. aureus* PVL-positive strains was <2% (Schaumburg *et al.*, 2014). However, MRSA as a prevalent pathogen in Middle East region has MRSA prevalence for 45% of *S. aureus* nasal colonised patients, most of these isolates were related to sequence type (ST) 22 and PVL-negative (Tokajian, 2014). Disruption of MRSA strains in Asian countries are similar to those observed in Europe and north American countries, to accounts for >50% of *S. aureus* isolates in some countries (Chen and Huang, 2014). *S. aureus* epidemiology in Australian indigenous communities indicated considerable changes in CA-MRSA clones over the past twenty years with high emergence and spreading rate, this spread was accompanied by a considerable increase in the prevalence of staphylococcal infections from 2000 to 2011(Williamson *et al.*, 2014).

Since *S. aureus* toxic shock syndrome (TSS) was first described by Todd *et al.* in 1978, TSS was linked with superabsorbent tampons in menstruating women (Herzer, 2001), with highest annual infection rate of 13.7 per 100,000 menstruating women (Osterholm and Forfang, 1982). Introduction of more hygienic advice to all susceptible women to use suitable tampons reduced the annual incidences of *S. aureus* TSS to 1 per 100,000 menstruating women and 0.3 per nonmenstruating women (Hajjeh *et al.*, 1999), after that, the incidence of *S. aureus* TSS has stayed stable with the annual incidences around 0.69 per 100,000 menstruating women and 0.32 per 100,000 of total populations (DeVries *et al.*, 2011).

Vaccine candidates for *S. aureus* have been tested in clinical trials using both active and passive immunization modalities, but in attempts to develop a vaccine in future there are several issues should be considered, firstly, antigens used should be expressed by a majority of *S. aureus* strains, second, these antigens have been proved to stimulate the immune response of preclinical animal models of infection, finally, in passive immunization animal experiments the produced antibodies against these antigens must be protective (Kuklin *et al.*, 2006). Four cell wall-anchored surface proteins of *S. aureus* as antigens in a murine model were assembled into a combined vaccine has given high levels of protection against invasive infections (Stranger-Jones *et al.*, 2006). The majority of antibacterial subunit-based vaccines components have two categories, either a secreted toxins or abundantly expressed surface exposed molecules (Grandi, 2010). IsdA and IsdH *S. aureus* surface proteins that are expressed during infection were used in cotton rat model of nasal colonisation, rats vaccinated with IsdA or IsdH developed protection against nasal carriage (Clarke *et al.*, 2006).

A number of excellent strategies to developing an effective *S. aureus* vaccine, including (1) a sensible fund needed to be invested in clinical trials. (2) A multiple mixed bacterial antigens will be preferable to become the main scientific approach for new vaccines. (3) Cell-mediated immune response biological function during infection should be evaluated in an appropriate patient population (Patti, 2011). An effective combined vaccine provided stable protection against *S. aureus* contains five conserved antigens known to have different roles in *S. aureus* pathogenesis including two lipoproteins, ferric hydroxamate binding lipoprotein FhuD2 and putative lipoprotein conserved staphylococcal antigen 1A (Csa1A) (Bagnoli *et al.*, 2015). It is very important to detect new vaccines and to find out new strategies to treat *S. aureus* infections.

1.1.5 Pathogenesis of *S. aureus*

Staphylococcus aureus is an adaptable microbe that can express a range of virulence factors, including adhesins, enzymes, toxins and capsular polysaccharides. These virulence factors are controlled by a set of staphylococcal regulatory networks, including eight main accessory gene regulators system (Bien *et al.*, 2011).

S. aureus is one of the pyogenic pathogenic bacteria with an ability to induce abscess in both local and metastatic infections. There are five stages in the pathogenesis of *S. aureus* infections; (1) colonisation, (2) local infection, (3) systemic dissemination and/or sepsis, (4) metastatic infection, and (5) toxinosis. Once the organism breach skin or mucous membranes and reach underlying host tissues, they can cause an illness in any part of the infected tissue, causing a range of diseases varying from minor skin infections to life-threatening systemic infections, such as endocarditis and haemolytic pneumonia. Adherence is the first step for bacterial colonisation of a new host, this process mediated by several adhesins. By expressing a range of surface bound proteins known as microbial surface components recognising adhesive matrix molecules (MSCRAMMs), these molecules recognize the most prominent components of the extracellular matrix and blood plasma including, collagen-binding protein, fibrinogen, fibronectin-binding proteins A and B, elastin binding protein, prothrombin binding protein and von Willebrand factor binding protein have been well characterised (Clarke and Foster, 2006, Foster and Höök, 1998). *S. aureus* has characteristic survival strategies during infection as shown in figure 1.

S. aureus produces various toxins that are classified on the basis of their mechanisms of action; it has three well known types of toxin: (i) cytotoxins, with main functions to lyse host cells to provide nutrients required for bacterial growth, (ii) leukotoxins (Luk), (iii) pyrogenic-toxin superantigens and (v) exfoliative toxins (Fueyo *et al.*, 2005). The pyrogenic-toxin superantigens are small sized proteins related and sharing some degrees of amino acid sequence homology, binding to major histocompatibility complex (MHC) class II proteins and produce an extensive T-cell proliferation and cytokine release (Marrack and Kappler, 1990). Different forms of enterotoxin molecule are responsible for the illness caused by these proteins such as toxic shock syndrome and food poisoning.

Even with limited amino acid sequence homology, toxic shock syndrome toxin 1 structure is similar to enterotoxins B and C, while the gene for toxic shock syndrome toxin 1 was present in 20% of *S. aureus* isolates (Marrack and Kappler, 1990). *S. aureus* produces a

range of exoproteins such as exotoxins and enzymes, the main function of these exoproteins may be to convert host tissue cells into nutrients required for bacterial growth (Dinges *et al.*, 2000). These cytolytic proteins form *β*-barrel pores in host cell plasma membrane and lyse the target cells cause leakage of cellular content (Foster, 2005). One of these important controversial virulence factor is PVL a cytotoxin bicomponent β-pore-forming toxin which produced by 2–3% of clinical *S. aureus* isolate (Kuehnert *et al.*, 2006), this factor was universally found in all CA-MRSA clones (David and Daum, 2010), clinical data linked the bicomponent cytolysin PVL with necrotising skin infections and pneumonia (Gillet *et al.*, 2002). Pathogenic strains differ in post-invasion strategies to give different disease signs, for example, the extremely pathogenic *S. aureus* strains 6850 and ST239 invade host cells and release different toxins and virulence factors regulated by agr global regulator system, while, other strains persist within intact cells without causing inflammatory affects and failed to express agr system virulence factors (Grundmeier *et al.*, 2010). Additional sets of exotoxins produced *S. aureus* which include the pyrogenic toxin superantigens (PTSAgs) toxic shock syndrome toxin-1 (TSST-1) and staphylococcal enterotoxins are able to stimulate proliferation of T-lymphocytes (Holtfreter and Broker, 2005). In toxin-mediated staphylococcal disease, since bacterial toxin (in this case one of several enterotoxins) has been elaborated, food poisoning can occur even in the absence of viable bacteria, however, in staphylococcal toxic shock syndrome (TSS) if toxin is elaborated at colonised sites, for example, in the presence of a superabsorbent tampons, this is sufficient to produce this syndrome (in this case menstrual TSS). Staphylococcal Scalded Skin Toxin SSS causes a skin disease mainly affecting children and immunocompromised patients. SSS toxin consists of two serotypes, exfoliatin A and B, which have superantigenic activity and therefore induce selective polyclonal expansion of T-cells restricted to certain Vβs T-cell receptors (Ladhani, 2003).

Infections with *S. aureus* occur often as a consequence of inoculation into an uncovered wound; however, in the upper respiratory tracts, viral infection damages mucosal layer and predisposes host to *S. aureus* pneumonia (McCullers, 2006). Early stages of *S. aureus* contact with host tissues without the mucosal layer or damaged skin triggers up-regulation of virulence genes (Novick, 2003). *S. aureus* main component peptidoglycan and lipoprotein are sensed by host pattern recognition molecules (Fournier and Philpott, 2005), also various endogenous Toll like receptor ligands released by necrotic tissues start to

trigger the pro-inflammatory signalling leading to different immune cell activation (Pisetsky, 2007).

A recent murine infection model study to determine the surface proteome of USA300 MRSA *S. aureus* found that the majority of *in vivo* expressed surface associated proteins were lipoproteins, which were part of ABC-type transport systems involved in nutrient acquisition especially metal ion uptake proteins (Diep *et al.*, 2014a).

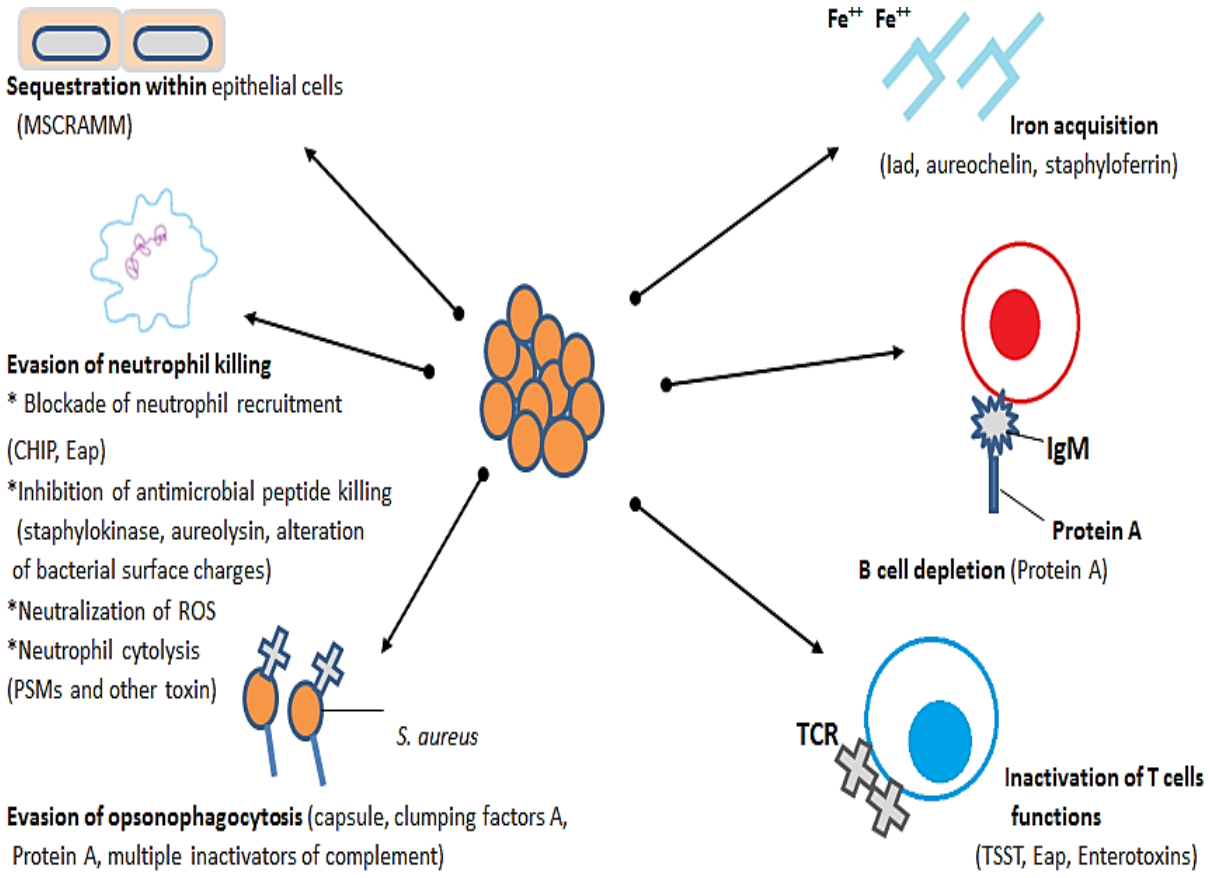

Figure 1. *S. aureus* survival strategies during infection stages
(adopted from Liu, 2004).

1.1.6 Detection of methicillin/oxacillin/cefoxitin resistance in *S. aureus* by using cefoxitin as test agent

Molecular and non-molecular methods are currently available to detect the presence of MRSA in clinical samples, correct detection of MRSA strains is very importance to ensure effective treatment and to prevent further transmission, methicillin/oxacillin-resistant staphylococci are heterogeneous in their showing a resistance to β-lactam antibiotics, while experimental conditions play a major role as well on the expression of these genes and altering the detection of resistance.

S. aureus susceptible and resistant strains produce four major penicillin-binding proteins (PBP 1, 2, 3 and 4) (Georgopapadakou and Liu, 1980), the fundamental cause of most methicillin resistance is production of an extra penicillin-binding protein PBP2a (PBP2') which determines the methicillin resistance strains (Hartman and Tomasz, 1984, Reynolds and Brown, 1985). PBP2a is a high molecular weight class B PBP (Ghuysen, 1994), PBP2a is located in the bacterial cell wall and has a low binding affinity for most of the semi-synthetic penicillins such as methicillin and oxacillin (Appelbaum, 2007), but the structural basis for this low affinity is not fully understood. PBPs proteins are mediated by the *mecA* gene, however, *mecA* homologous genes with 80% nucleotide similarity were found in *Staphylococcus sciuri* (Hanssen and Ericson Sollid, 2006), also homologous genes with 91% nucleotide similarity were detected in *Staphylococcus vitulinus* (Schnellmann *et al*., 2006). There are some other genes that may affect expression of methicillin resistance in *S. aureus* but these genes were detected in both susceptible and resistant strains (Labischinskia *et al.*, 1998). Also some strains appear to have a low level of resistance with variations to existing of PBPs (Bignardi *et al.*, 1996). Strains that produce extra penicillinase have shown low level of resistance in some conditions (Mcdougal and Thornsberry, 1986). Cross resistance between methicillin and other β-lactam antibiotics was identified in MRSA strains (Chambers, 1997).

Cefoxitin disk diffusion test has been used to predict the presence of *mecA*-mediated oxacillin resistance in *S. aureus* strains which encodes PBPs that related to oxacillin resistance of Coagulase-negative staphylococci CoNS (Chambers, 1997). Cefoxitin disk diffusion test results are easier to interpret and more sensitive for the detection of *mecA*-mediated resistance strains than oxacillin disks (Pottumarthy *et al.*, 2005, Witte *et al.*, 2007a). In 2007, the Clinical and Laboratory Standards Institute (CLSI) modified the

recommended resistance and susceptibility breakpoints for the 30 μg cefoxitin disk test to detect *mecA*-mediated resistance in *S. aureus* from ≤19 mm and ≥20 mm to ≤ 21 mm and ≥ 22 mm, respectively (Broekema *et al.*, 2009). The international guidelines for using oxacillin to define methicillin resistance strains as resistance ≥4 mg/L and susceptible ≤ 2 mg/L (Howe and Andrews, 2012).

1.1.7 Cell envelope proteins of staphylococcal and their functions

Staphylococci have a number of unique apparatuses to immobilise proteins on their surface, either via covalently linked by their C-terminal to cell wall peptidoglycan or with non-covalent binding of proteins to either the peptidoglycan or secondary wall polymers e.g. teichoic acids (Navarre and Schneewind, 1999). These proteins play a role in bacterial pathogenicity by establishing successful colonisation, invasion of host tissue and surviving in the host environment including; adhesion, antiphagocytic influence, destruction of host cell surface components and hydrolysis of molecules for nutrient utilisation (Lee and Fischetti, 2006). Some of these proteins are shown in figure 2. Staphylococci have no pili or fimbrial structures and employ surface protein-mediated adhesion to host cells as a mechanism to escape from immune defenses also to survive within the infected host (Telford *et al.*, 2006). Many cell wall proteins are also responsible to organise synthesis and maturation of bacterial peptidoglycan at particular sites during cell growth and division (Höltje, 1998).

Surface proteins of staphylococci that are anchored in the cell wall surface consist of at least two topogenic sequences, i.e., an N-terminal signal peptide and C-terminal cell wall sorting signal (Abrahmsen *et al.*, 1985). Sortases enzymes promote the covalent anchoring of surface proteins to the cell wall envelope, this enzyme catalyse transpeptidation reaction within the first cleaving surface protein substrate at the cell wall sorting signal, the outcome of this reaction a surface protein linked to peptidoglycan which combined into cell envelope and displayed on the bacterial surface (Schneewind *et al.*, 1993). Generally, all bacterial proteins destined for translocation across the cytoplasmic membrane consist of an N-terminal signal recognition sequence, but through the translocation process signal peptides are removed proteolytically via signal peptidases then these proteins are either secreted into the extracellular matrix or attached to the cell wall (Dalbey *et al.*, 1997). Covalently linked proteins to peptidoglycan layer possess a C-terminal sorting signal containing a conserved LPXTG sequence motif (Navarre and Schneewind, 1999). Because

of their extracellular localization another class of *in-vivo* expressed surface-associated proteins of *S. aureus* are lipoproteins (Diep *et al.*, 2014a, Sutcliffe and Harrington, 2002). A new bioinformatics experimental tool can be used to predict bacterial proteins exposed on the surface of the organism that commonly involved in host-pathogen interaction to identify antibacterial targets for both therapy and vaccination (Giombini *et al.*, 2010).

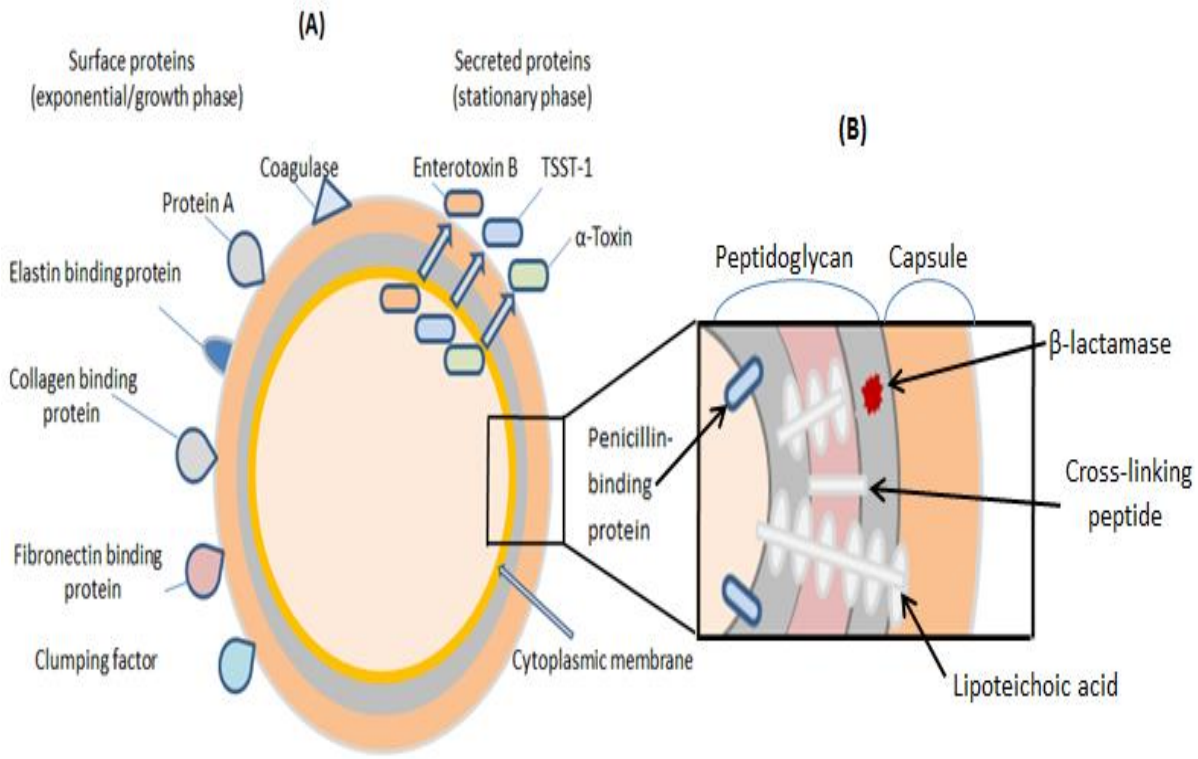

Figure 2. (A) *S. aureus* cell surface and secreted proteins in a different bacterial growth phases. (B) Shows a cross section of the bacterial cell envelope
(Lowy, 1998)

1.1.8 LPXTG motif (Leu-Pro-any-Thr-Gly)

Cell wall associated surface proteins have many common structures to enable them to anchor to the cell wall peptidoglycan. These features consists an N-terminal signal peptide required for Sec pathway secretion and a conserved the C-terminal cell wall sorting signal that helps in attachment of a protein to the cell wall by sortase (SrtA) (Mazmanian *et al.*, 1999). One prevalent feature of peptidoglycan linked surface proteins is the presence of the LPXTG-motif at the C-terminal, during surface translocation. For example, protein A of *S. aureus* is produced as a precursor bearing an N-terminal signal peptide and a C-terminal sorting signal with an invariant LPXTG motif, attached to a series of hydrophobic amino acids and a short tail of positively charged residues (Mazmanian *et al.*, 2001).

A specific transpeptidase enzyme srtA cleaves the peptide bond between the threonine and glycine residues of LPXTG motif, then covalently links them to the peptidoglycan via the carboxyl group of threonine, subsequently tethering the C-terminus of protein A to the bacterial peptidoglycan (Jonsson *et al.*, 2002). Mutant *S. aureus* strains lacking srtA gene failed to anchor and display some surface proteins and showed less ability to cause infections (Mazmanian *et al.*, 2000). *S. aureus* surface has been reported to display up to 21 different LPXTG proteins (Roche *et al.*, 2003). However, according to an online database on (http://www.uniprot.org), which provides the scientific community with a comprehensive high quality and freely accessible resource of protein sequence and functional information, *S. aureus* has about 32 LPXTG proteins after analyses of 14 complete genome sequences, whereas only 21 genes encoding LPXTG proteins were originally identified (Roche *et al.*, 2003). LPXTG include the surface proteins, Bap (biofilm associated protein) and Aap/SasG (*S. epidermidis* accumulation-associated protein/ *S. aureus* surface protein G), these proteins are mediators of staphylococcal biofilm development (O'Neill *et al.*, 2008). *S. aureus* mutant strains, lacking the *srtA* gene, failed to anchor all surface proteins examined due to a defect in processing of sorting signals at the LPXTG motif (Mazmanian *et al.*, 2000). Most of these polypeptides were found to react with human tissues, serum proteins and polypeptides of the extracellular matrix (Foster and Höök, 1998).

1.1.9 Host responses to *S. aureus* infection

Different bacterial subcellular components have been reported to activate the host cells response, such as bacterial envelope elements e.g. lipoproteins, peptidoglycan, teichoic acids or secreted compounds such as enterotoxins or toxic shock syndrome toxin (Stoll *et al.*, 2005). Invading bacteria replicate in infected tissues and induce proinflammatory responses with the release of cytokines and chemokines which are necessary for recruiting immune cells to the site of infection. Invasions of host immune cells occurs together with partial liquefaction and necrosis of tissue and production of peripheral fibrin walls to avoid microbial spread and preparation to eliminate the necrotic tissue (Jonsson *et al.*, 1985).

Detection of the invading pathogen is the first step of host immune system, the innate immune response is capable of recognising pathogens and provides a relative first line of defence. This recognition requires pattern recognition receptors (PRRs), including Toll-like receptors (TLRs) and intracellular nucleotide-binding oligomerization domain receptors (NLRs) (Akira *et al.*, 2006). *S. aureus* virulence is related to different bacterial surface components (e.g., polysaccharide capsule and protein A), as well as surface bound proteins for example, clumping factor and fibronectin binding proteins and also extracellular proteins (e.g. coagulase, hemolysins, enterotoxins, TSST-1, exfoliatins and Panton-Valentine leukocidin (PVL) (Archer, 1998).

The ability of bacteria to cause disease is due to evasion of host defence; this includes resistance to antimicrobial peptides and killing by phagocytic cells (Levy, 1996). In the lungs, defences include mucociliary clearance of the respiratory system epithelium and increasing the production of antimicrobial peptides, surfactant proteins, chemokines and cytokines mediating immune cells which help to prevent colonisation by pathogens (Bals and Hiemstra, 2004). Host epithelial cells forms the initial line of defence against microbial pathogens via activation of various intracellular signalling pathways, *S. aureus* evokes an intense host response to trigger the poly morphonuclear leukocyte (PMNs) cells to the site of infection by chemotactic signaling pathways, these immense numbers of leukocytes are considered the primary cellular defence against *S. aureus* invasion and the subsequent infiltration of macrophages and fibroblasts (DeLeo *et al.*, 2009). Bacterial Lipoproteins (Lpp) as a pathogen-associated molecular patterns (PAMP's) are sensed by host immune cells by TLR2 (Toll-like receptor 2) together with TLR1 or TLR6, depending on this signals the innate immunity will induce and activate the necessary pathways to

control adaptive immunity (Iwasaki and Medzhitov, 2004). Eleven human TLRs and 13 mouse TLRs have been identified so far, each TLR is able to identify pathogen associated molecular patterns derived from different microbes (Akira *et al.*, 2006). Bacterial lipoprotein act as trigger molecules to activate the host innate immune responses via TLR2 and subsequently TLR signals participate in direct regulation of adaptive immunity (Iwasaki and Medzhitov, 2010). Synthetic lipoprotein analogs such as Pam3Cys lipopeptides, Pam3CSK4 from *Escherichia coli* and dipalmitoyl MALP-2 from *Mycoplasma fermentans* have shown similar proinflammatory properties of bacterial lipoproteins (Takeuchi *et al.*, 2000). Further experiments led to clear evidence in which triacylated lipopeptides recognise through TLR2/TLR1, while diacylated lipopeptides recognise through TLR2/TLR6 (Akira, 2003).

Lipoproteins are the dominant immunobiologically active compound in *S. aureus* by their influence to induce cytokine release (Hashimoto *et al.*, 2006b). An evidence shows that released lipoprotein were important for induction of pro-inflammatory cytokine interleukin-1β (IL-1β) and the activation of Nlrp3 inflammasome the member of intracellular nucleotide-binding oligomerization domain receptors (NLRs) and involved in immune defence against infection (Muñoz-Planillo *et al.*, 2009). Impact of lipoprotein on the inflammation *in vitro* and immunocompetent cases *in vivo* was dependent on mature lipoprotein through releasing a strong cytokine and chemokine response by activating TLR2 (Schmaler *et al.*, 2009). Meanwhile, lipoprotein deficient *S. aureus* RN4220 strain did not induce TLR2 activation (Kim *et al.*, 2015). Lipoproteins were able to stimulate the activity of various immune cells types via TLR2 to release cytokines and chemokines products, such as monocytes and osteoclasts (Kang *et al.*, 2011, Kim *et al.*, 2013b). *S. aureus* deleted νSaα specific lipoprotein like cluster genes (*lpl*) showed that mutant bacterium was lacking in the stimulation of pro-inflammatory cytokines of host monocytes, macrophages and keratinocytes (Nguyen *et al.*, 2015).

1.2 Lipoproteins

1.2.1 Lipoprotein biosynthesis and localisation

In general, cellular proteins can exist in different forms, either in soluble form in the cellular spaces (cytoplasm in both monoderm and diderm bacteria or periplasm in diderms only), or anchored to cell membranes (cytoplasm membrane in monoderms, inner- or outer membrane in diderms), some are anchored to cell wall (in monoderms), whereas, other proteins can be translocated into host cells or released into the extracellular spaces (Desvaux *et al.*, 2009). The attachment of lipids to cellular proteins is an important post-translational modification occurring in both prokaryotes and eukaryotes. In 1969, some modifications were discovered in the major outer membrane protein of *E. coli* within the lipid N-acyl-S-diacylglyceryl cysteine at the N-terminal (Braun and Rehn, 1969). Afterwards, many related and unrelated bacterial proteins with the same lipid modification were discovered and are generally known as lipoproteins. The attached lipid to the outer membrane protein was later known as a diacylglyceryl group, this moiety attached by thioether linkage to the sulfhydryl group of N-terminal cysteine and the α-amino group of diacylglyceryl modified cysteine is fatty acylated as in figure 3 (Hantke and Braun, 1973). Another study in the Gram-positive bacterium *Acholeplasma laidlawii* have established a major class of membrane lipoproteins which share type II signal peptide sequences with a conserved lipid-modified cysteine residue at the N-terminus to enable this protein to anchor onto the periplasmic leaflet of the plasma membrane or outer membrane in Gram-negative bacteria (Dahl *et al.*, 1985). Bioinformatic studies on the available bacterial genomes sequences indicated that lipoprotein genes constitute approximately 1–3% of their total genes (Babu *et al.*, 2006).

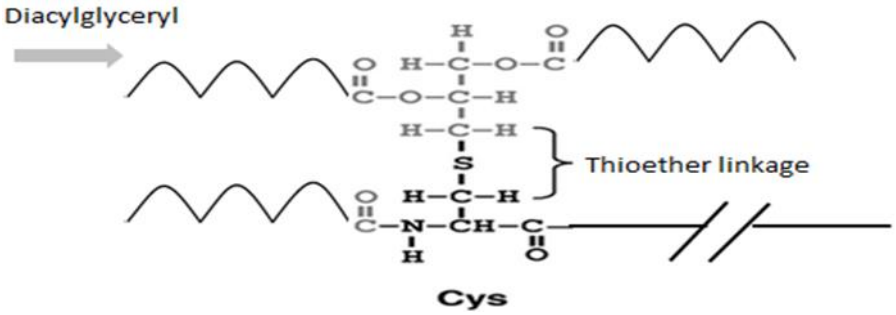

Figure 3. The structure of the lipid modification in lipoproteins
The N-terminal cysteine modified with a diacylglyceryl group attached by a thioether linkage, the amino group is acylated with a fatty acid.

Bacterial cellular activities required a diverse class of membrane proteins that can work well in aqueous medium, whilst anchored to the hydrophobic membrane of the cell envelope. Bacteria have evolved several strategies for their various membrane proteins: e.g. (1) proteins with a hydrophobic surface, which along with other noncovalent and even ionic interactions, associate with the membrane; (2) transmembrane proteins carrying peptide segments in their helical or beta sheeted structure cross the membrane to provide anchorage and help the parts of the transmembrane segments perform the relevant roles; (3) some proteins have lipid modification with exo or endo fatty acids, as well as other lipid moieties, that provide a hydrophobic anchor either at one end or on the surface of such proteins (Babu *et al.*, 2006). Bacterial lipoproteins have been classified according to their functional nature as antigens, adhesins, binding proteins, enzymes, transporters, toxin, structural proteins and hypothetical lipoproteins (Babu and Sankaran, 2002).

Cell envelope lipoproteins of Gram-positive bacteria are anchored into the outer leaflet of the plasma membranes, a lipid modification takes place by covalent addition of a diacylglyceride to an indispensable cysteine residue located in the C-terminal region of a signal peptide as described for the prototypical Braun's lipoprotein of *E. coli* (Braun and Wu, 1994). lipoprotein are effective within a subcellular part located between the inner aspect of plasma membrane and outer aspect of the peptidoglycan and other layers of the cell wall (Hutchings *et al.*, 2009). In the absence of an outer membrane in Gram-positive bacteria, proteins must be attached to the plasma membrane to be retained within the cell envelope. For this reason, several lipoprotein have functions similar to the periplasmic or surface proteins of Gram-negative bacteria (Rahman *et al.*, 2008).

Consequently, it is suggested that at least some lipoproteins of Gram-positive bacteria are functionally similar to periplasmic proteins of Gram-negative bacteria, a comparison most directly sustained by the fact that, in Gram-positive bacteria, substrate binding proteins (SBPs) of ATP-binding cassette (ABC) transporters are classically lipoproteins (Sutcliffe and Russell, 1995). Functionally, many bacterial lipoprotein have a role in substrate binding action, which is a part of ABC transport systems (Tam and Saier, 1993), these proteins are also involved in sensing environmental signals (Sutcliffe and Russell, 1995), protein secretion and folding of exoproteins (Kontinen and Sarvas, 1993), adherence to various surfaces (Kolenbrander *et al.*, 1998), invasion of host cells (Réglier-Poupet *et al.*, 2003), bacterial coaggregation (Kolenbrander, 1993), antibiotic resistance (Jousselin *et al.*, 2012), respiration (Bengtsson *et al.*, 1999). A little is known about the biochemical

functions of individual lipoprotein but their proteomics has not been investigated as much in Gram-positive bacteria as they have in Gram-negative bacteria. Determination of the accurate lipidated structures of lipoprotein is crucial for elucidating the molecular basis of interactions between the host and microorganism.

1.2.2 Lipobox of *S. aureus* lipoproteins and their modification

Most bacterial proteins that are synthesised within the cell and transferred to extra cytoplasmic space or growth medium have an N-terminal signal peptide or signal sequence. Analysis of the signal sequences of several lipoproteins revealed common structural features that are recognised prior to lipid modification. The signal sequence is divided into three regions: secretory signal peptides structures contain a short positively charged N-region, a hydrophobic H-region that spans the membrane and a C-region has small and uncharged residues around the cleavage site which is recognised by the peptidase to cleave the peptide and produce a mature protein (Heijne, 1983). Signal peptides are mainly divided into secretory signal peptides that are cleaved by Signal Peptidase I and others cleaved by Signal Peptidase II, which is characteristic of the membrane-bound lipoproteins (Sankaran and Wu, 1995). Lipoproteins are initially translated as preprolipoproteins in both Gram-negative and Gram-positive bacteria, which have an N-terminal signal peptide, containing ~ 20 amino acids with distinctive characteristic elements of the signal peptides of secreted proteins (Inouye *et al.*, 1977). The C-region of lipoprotein signal peptides contain a four-amino-acid motif called the lipobox (Sankaran and Wu, 1993). Proteins intended to be lipidated carry a conserved sequence at the C-region end of signal peptides with Cysteine (+1 position) to which the lipid-modification linked is invariant and the (-3 position) mainly Leu, referred to as the lipobox, with consensus of lipobox amino acid sequence of [LVI] [ASTVI][GAS]C, this directs them to the lipoprotein biogenesis maturation. The N-region contains from 5 to 7 residues with two positively charged Lys or Arg residues and the H-region length varies between 7 and 22 residues, with a modal value of 12 residues (figure 4). The C-region motif forms the molecular basis for many in silico algorithms that have been used to predict lipoprotein genes in bacterial genomes (Babu *et al.*, 2006, Setubal *et al.*, 2006).

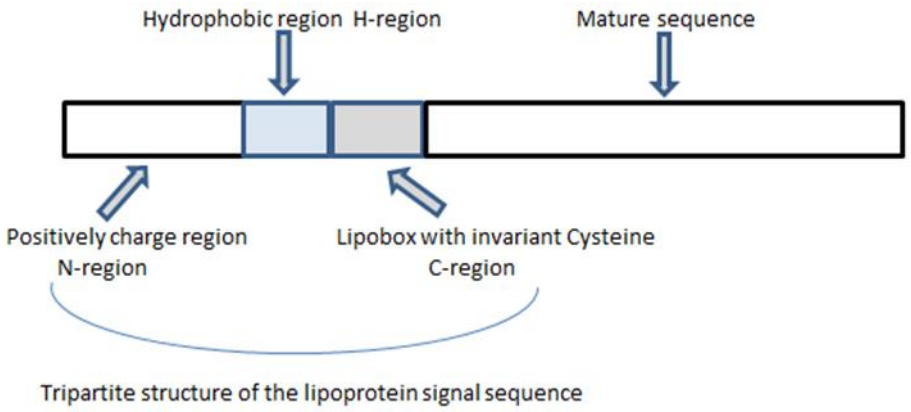

Figure 4. Tripartite structure of lipoprotein signal sequence
Modified from Babu *et al*. (2006).

The sequence of the C-region of signal peptides (lipobox) with sequences [LVI] [ASTVI][GAS]C, is modified through the covalent attachment of a diacylglycerol moiety to the thiol group on the side chain of the essential cysteine residue, including a regular four amino-acid sequence at the C-terminal end of signal peptide sequence –Leu_3– Ser/Ala_2–Ala/Gly_1–Cys_1 ending with the modifiable cysteine (Babu and Sankaran, 2002). A statistical survey on amino acids of lipobox predicted lipoproteins from 234 completely sequence of different bacterial genomes preceding the lipid-modifiable Cys (+1 position) shows that these are well conserved, about 70% of the models showed that the -3 position was Leu (71%), Val (9%) and Ile (6%). Also, occasionally Ala, Phe, Gly, Cys, or Met residues were found in the -3 position but with low frequencies (<5%). The -2 position was more variable and could contain uncharged polar and nonpolar residues Ala (30%), Ser (28%), Thr (12%), Val (10%) and Ile (8%). Gly, Leu and Met were observed with low frequencies in this position, usually, the -1 position was occupied equally by Gly (45%) or Ala (39%); but Ser was found in 16% of all results (Babu *et al.*, 2006). Predictions of lipoprotein *in silico* and structure information indicate that the amino acid following +1 cysteine lack any predicted or observed secondary structure (Zückert, 2014).

1.2.3 Post-translational lipid modification of lipoproteins

Sec-associated YidC appears to assist preprolipoproteins pass the cytoplasmic membrane as unfolded proteins via the general secretory (Sec) pathway (Fröderberg *et al.*, 2004), or with the help of a SecA1 and SecA2 (Feltcher *et al.*, 2013), however, other secretion pathways are used to transport lipoprotein across the membrane in some high GC Gram-positive bacteria via the twin-arginine translocation pathway (Sheldon and Heinrichs, 2012). The biosynthetic pathway of bacterial lipoprotein involves three sequentially acting enzymes: preprolipoprotein diacylglyceryl transferase (Lgt), prolipoprotein signal peptidase (Lsp) and apolipoprotein N-acyltransferase (Lnt). The first step of biosynthetic pathway involves the lipoprotein diacylglyceryl transferase (Lgt), which attaches the diacylglyceryl group from phosphatidylglycerol to the thiol of Cys, (first amino acid after the signal peptide) via a thioether linkage resulting in a prolipoprotein. The second enzyme Lsp (lipoprotein signal peptidase), recognises the diacylglyceryl modification and cleaves between the amino acid at position -1 and the lipid-modified cysteine residue, and leaves cysteine of the lipobox shown as new amino-terminal residue (Tokunaga *et al.*, 1982). Lipoproteins from Gram-negative bacteria are further modified by lipoprotein N acyltransferase (Lnt) which adds an N-acyl group to the diacylglyceryl cysteine and produces a mature triacylated lipoprotein triacylated lipoprotein (Sankaran and Wu, 1994). These enzymes are important in lipoprotein biosynthesis in Gram-negative bacteria, but, in Gram-positive bacteria *Lgt* and *Lsp* have been found to be essential in some of the high GC-content species tested, but not in low GC-content species such as *S. aureus* (Nakayama *et al.*, 2012). Bacterial lipoproteins are structurally split into two types; diacylated lipoproteins and triacylated lipoproteins, this classification depends on the absence or presence of particular enzymes involved in lipoprotein maturation as shown in figure 5.

N-acylation modification in *Escherichia coli* is essential for sorting lipoproteins from the inner membrane to the outer membrane by localization of lipoprotein via (Lol) system (Tanaka *et al.*, 2001). *E. coli* type Lnt enzyme is conserved in most Gram-negatives and in high GC-content Gram-positive bacteria, has been shown to contain triacyl lipoproteins (Thompson *et al.*, 2010). *Lnt* homolog was not found in the genome of some *S. aureus* strains suggesting that lipoproteins are diacylated (Stoll *et al.*, 2005). However, Kurokawa *et al*. reported a purified triacylated 33kDa lipoprotein *S. aureus* was able to stimulate TLR2 and was confirmed as a triacylated SitC lipoprotein (Kurokawa *et al.*, 2009). In addition, genomic analysis has revealed that some members of high-GC-content Gram-

positive bacteria (actinomycetes) contain homologues of Lnt the final enzyme involved in the maturation of a lipoprotein (Vidal-Ingigliardi *et al.*, 2007). Triacylated lipoproteins were detected in low-GC Gram-positive *S. aureus* (Asanuma *et al.*, 2011, Kurokawa *et al.*, 2009). Also, putative Lnt genes were identified in Gram-positive mycobacteria (Tschumi *et al.*, 2009). Recent biochemical evidence that N-acylation-free diacyl lipoprotein accumulated when bacteria were grown in acidic or high salt concentration media (Kurokawa. *et al.*, 2012), all these results suggest that bacterial lipoprotein biosynthesis is changeable in response to growth conditions and variable between bacterial strains.

Both Lgt and Lsp enzymes are extensively conserved in eubacteria, but Lnt has not been confirmed in all low G-C content Gram-positive bacteria (Asanuma *et al.*, 2011, Stoll *et al.*, 2005). Consequently, whether lipoproteins of *S. aureus* are di- or triacylated is still not confirmed. Lipoprotein in low GC-content bacteria such as Firmicutes and Tenericutes had been considered as diacylated with the absence of *E. coli* Lnt gene in their genome (Kovacs-Simon *et al.*, 2011, Tschumi *et al.*, 2009, Zückert, 2014), and lipoprotein in Tenericutes was shown to have diacyl form (Shibata *et al.*, 2000), also contained the N-acylated triacyl form (Serebryakova *et al.*, 2011). MS/MS analysis on lipoprotein structures of two related low G-C Gram-positive bacteria *M. genitalium* and *M. pneumoniae* were found in N-acylated triacyl form (Kurokawa *et al.*, 2012), these uneven results predict that unidentified Lnt similar enzyme are involved in the biosynthetic pathways of the N-acylated triacyl and N-acetyl lipoprotein forms.

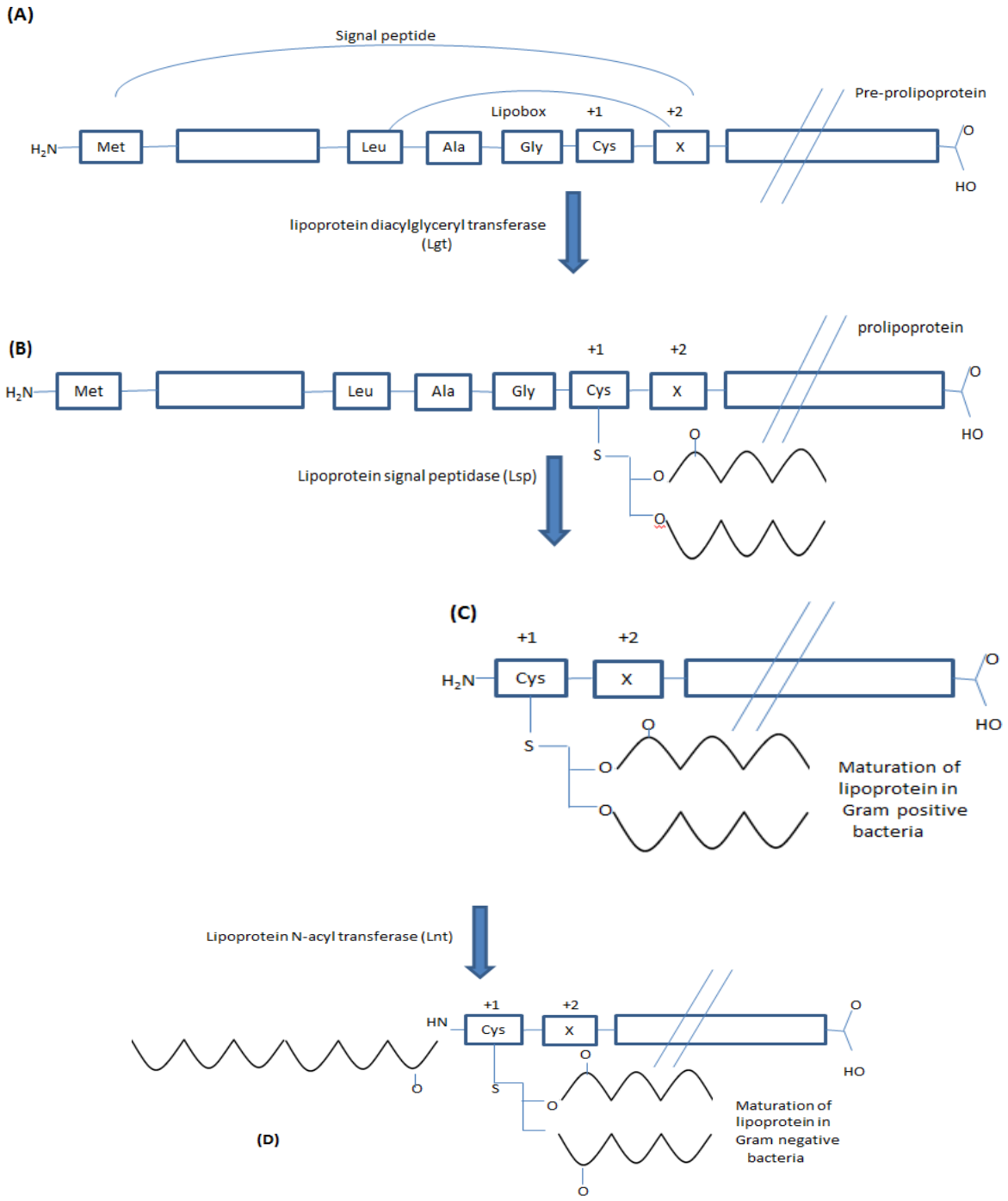

Figure 5. Biosynthesis pathway of bacterial lipoprotein in Gram-positive (two-step) and (A to D) biosynthesis pathway of Gram-negative bacteria (three-step)

(A) The precursor of lipoprotein is preprolipoprotein translocated by the Sec or Tat machinery into the outer leaflet of the plasma membrane, (B) thiol group of invariant cysteine in lipobox is modified by a diacylglyceryl moiety by Lpp Lgt transfers a diacylglyceryl, generating a thioether linkage. (C) Lsp cleaves the signal peptide at N-terminus leaving the cysteine as new amino-terminal residue forming the mature lipoprotein in Gram-positive. (D) In Gram-negative and some Gram-positive bacteria Lnt enzyme, transfers an acyl group from another phospholipid to the newly-generated a-amino group of the S-diacylglyceryl cysteine of the apolipoprotein, to form mature triacylated lipoprotein. Modified from Kovacs-Simon *et al*., 2011.

Preprolipoproteins are transported across the cytoplasmic membrane by one of two different pathways. Firstly; the general secretory pathway (Sec), which is the predominant route of protein transport for those proteins carrying a secretory signal peptide by the action of Sec translocase which recognises proteins bearing N-terminal signal peptides and transfers them across the membrane in an unfolded conformation (Driessen and Nouwen, 2008). The Sec pathway exports unfolded proteins using the energy generated by ATP hydrolysis (Natale *et al.,* 2008). The majority of preprolipoproteins in Gram-negative *Escherichia coli* are exported by Sec machinery (Sugai and Wu, 1992), Sec was shown to influence the Lsp mediated mechanism of lipid modified prolipoprotein that are formed by Lgt (Kosic *et al.*, 1993).

Secondly; via the twin arginine protein translocase pathway (Tat), this recognises signal peptides that carry a distinctive form of two consecutive Arginines (R-R) in the N-region (Lee *et al.,* 2006). Tat translocase exports folded proteins possessing a twin-arginine motif and is reliant on the proton motive force (Natale *et al.,* 2008). Translocation of lipoproteins by the (Tat) system in the high-GC Gram-positive bacteria Actinomycetes has been demonstrated (Kovacs-Simon *et al.,* 2011). Also further modification has been reported on lipoprotein for attachment to the cell wall or anchoring in the cytoplasmic membrane (Schmaler *et al.*, 2010). Bioinformatic analysis study in high-GC-content Gram-positive *Streptomyces* species has indicated that a significant fraction of preprolipoproteins were exported via Tat with up to 20% of putative lipoprotein (Zückert, 2014). Meanwhile no information are available on Tat dependent lipoproteins in the low-GC-content Gram positive bacteria, even some Firmicutes genomes shown lack of Tat pathway (Dilks *et al.*, 2003), it has not been proved that lipoproteins can be transported via the Tat pathway, therefore mature lipoproteins in the membrane appear to be exported via the Sec pathway.

1.2.4 Roles of lipoproteins in bacterial pathogenesis

The significance of lipoproteins from the point of their roles in bacterial pathogenesis is of interest, as these lipid-modified proteins play a variety of roles in host-pathogen interactions. *S. aureus Lgt* mutant strains were measured for production of proinflammatory cytokines and found to produce lower levels of TNF-α and IL-6 than wild-type strains (Wardenburg *et al.*, 2006). In the same manner, *Lsp* mutant strains of *Listeria monocytogenes* were found to be ineffective in phagosomal escape of bacteria during invasion (Réglier-Poupet *et al.*, 2003). The antibiotic globomycin acts as a specific

inhibitor to SPase II; bacteria were treated with globomycin and SPase II deficient mutant strains showed accumulation of lipid-modified prolipoproteins (Hayashi and Wu, 1990). Also, lipoproteins released from gram-negative Enterobacteriaceae played a role in activating the inflammatory response and avoiding the host defence by inducing cytokine production in the macrophage (Zhang *et al.*, 1998). Moreover, a 19 kDa lipoprotein of Mycobacteria that elicits antibody and T cell responses in humans and mice, induced innate immune response in dendritic cells and neutrophils (Neufert *et al.*, 2001). Modified antigens of lipoproteins are good candidates in vaccine development, for instance, lipoprotein 20 is an outer membrane lipoprotein that is an excellent vaccine candidate antigen against *Helicobacter pylori* (Keenan *et al.*, 2000), while the vaccine of Lyme disease was generated from OspA and DbpA lipoproteins of the spirochete *Borrelia burgdorferi*, which was found to be effective in several animal models (Chang *et al.*, 1995, Hanson *et al.*, 1998). Expression of PPIase PrsA lipoprotein in *S. aureus* affects both glycopeptide and oxacillin resistance in MSSA and MRSA strains (Jousselin *et al.*, 2012). Five lipoproteins were involved in nutrient acquisition in murine identified model (Diep *et al.*, 2014a). Lipoprotein FhuD2 immunisations has shown protective immunity against *S. aureus* in murine infection model and proved as an effective vaccine candidate (Mishra *et al.*, 2012). Mariotti, *et al*. confirmed the efficiency of apo-FhuD2 as a protective antigen, vaccination with FhuD2 or FhuD2 formulated with hydroxamate siderophores were protective in a murine *S. aureus* infection model (Mariotti *et al.*, 2013).

1.3 *S. aureus* genome

Staphylococcal genomes are approximately 2.8 Mbp in size with relatively low G-C content. *S. aureus* genome sequencing was an essential step for future development against this pathogen. The first sequenced *S. aureus* strains N315 and Mu50 (Kuroda *et al.*, 2001) were determined by shot gun random sequencing and this has been followed by many other additional strains (Bartels *et al.*, 2014, Nair *et al.*, 2011, Schijffelen *et al.*, 2013). Identification of genes associated with pathogenic strains would help in investigation of *S. aureus* variation, evolution, pathogenicity and epidemiology. Hence, to understand the roles of ecological and genetic features, especially host-pathogen molecular interactions, involved in host-to-host transmission and colonisation, it is important to expose new strategies to control the pathogen. A very complicated regulatory network is responsible for many differential gene expressions. *S. aureus* carry genes encoding toxins, cell surface

proteins and antimicrobial resistance that may increase its virulence and give resistance to all antibiotics families; expression of all these genes is regulated by specific and very sensitive mechanisms (Fitzgerald *et al.*, 2001). Different methods such as: (i) multilocus enzyme electrophoresis (MLEE); (ii) random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR); (iii) restriction fragment length polymorphism (RFLP); (iv) pulsed-field gel electrophoresis (PFGE); (v) staphylococcal protein A (*spa*) typing ;and (vi) multi-locus sequence typing (MLST) have been used to investigate the molecular epidemiology of staphylococci (Ben Nejma *et al.*, 2014, Qu *et al.*, 2014).

The completed genome sequences of numerous strains of *S. aureus* suggested that there was: (i) a high degree of nucleotide sequence similarity among the different strains; (ii) acquisition of genetic information by horizontal transfer between bacterial species had occurred and (iii) there was a unique number of pathogenicity or genomic islands and mobile genetic elements that contain clusters of enterotoxin and exotoxin genes or antimicrobial resistance determinants (Baba *et al.*, 2008). The first microarray comparative genomics studies to characterise *S. aureus* genomic diversity and virulence gene distribution among 36 strains covered 92% of genes in *S. aureus* COL genome, identified that 22% of *S. aureus* compared genomes was variable (Fitzgerald *et al.*, 2001).

Whole genome comparative analytical study of 12 *S. aureus* strains revealed unique features, among all of the sequenced strains 32.8 to 33% G-C contents did not vary significantly, with chromosome lengths range from 2.74 to 2.91 Mbp, genomic islands found included prophages (range of 1 to 4) and pathogenicity islands (PI) were found in different numbers (Baba *et al.*, 2008). Genome comparative studies on *S. aureus* revealed that the genome was composed of a complex combination of genes, many of them have been acquired by lateral genetic transfer (LGT) determine virulence important properties and antimicrobial resistance, but insufficient information is known about transfer mechanisms (Hiramatsu *et al.*, 2004). The common and unique genomic island νSaα in *S. aureus* encode tandem paralog referred to as lipoproteins cluster (lipoprotein-like [*lpl*]) comprises 10 lipoprotein genes with a lipo-box containing signal sequence (Babu *et al*., 2006). The νSaα mobile genetic elements suggested to be acquired independently through intra-species genetic transfer between *S. aureus* strains (Baba *et al*., 2008).

Lindsay *et al*.(2006) used multi-strains whole genome microarrays to compare 61 community-acquired invasive isolates of *S. aureus* and 100 carriage isolates from healthy

individuals, 10 human lineages were found to dominate and some minor lineages, each lineage possessed a unique combination of surface proteins and regulators called core-variable (CV) genes, which match to the clonal complexes (CCs) specified by MLST typing, this lineage has not evolved independently, but as a consequence of multiple recombinations of CV genes between them, also, all strains hold a range of mobile genetic elements (MGE) and make up approximately 15% to 20% of the genome and include bacteriophages, integrated plasmids, pathogenicity islands, staphylococcal cassette chromosomes (SCC), genomic islands and transposons (Lindsay *et al.*, 2006). All of these emerge to play a role as putative virulence genes and resistance, but variations within lineages indicate regular horizontal transfer (Witney *et al.*, 2005). A few extant clones are linked with the majority of infections and some of lineages are non-randomly related with distinct human and animal infections, whether a limited number of lineages are responsible for a large part of *S. aureus* infections implies that interclone variance in relative virulence is large and raises the possibility that difference in genome content can lead to significant roles in pathogenicity (Fitzgerald *et al.*, 2001).

A total of 56 UK isolates of animal associated *S. aureus* and 161 human *S. aureus* isolates from healthy individuals and community acquired infections in UK, determine that animal-associated *S. aureus* group was related to ten lineages, with approximately 60% of them assigned to only four lineages, most mastitis cases were caused by bovine strains, however a few human lineages caused mastitis, nearly 54% of horse-associated *S. aureus* isolates belonged to human associated lineages clusters (Sung *et al.*, 2008).

S. aureus has many well characterised global regulators of virulence determinant production; such as accessory gene regulator *agr* (Peng *et al.*, 1988), staphylococcal accessory regulator *sarA* (Cheung and Projan, 1994), *sae* (Giraudo *et al.*, 1994), *sigB* (Bischoff *et al.*, 2001), *arl* (Fournier and Hooper, 2000), and some Sar proteins repress one or more of *sarA* homologues genes (Arvidson and Tegmark, 2001), each of these regulators are parts of an important network involved in control and expression of virulence factors including surface proteins expression, exoproteins and many other proteins essential for bacterial activity (Harris *et al.*, 2002). One virulence gene can be under the control of many regulators to ensure that the specific gene is expressed under favourable conditions. For example, agr regulator negatively regulates the expression of *spa*, encoding protein A (Projan and Novick, 1997), although *SarS* activates its expression via binding to spa promoter (Tegmark *et al.*, 2000). In the same time, agr down-regulates

sarS expression level (Cheung *et al.*, 2001). This indication suggests that agr down-regulates spa expression when down-regulating expression of *sarS* activator (Tegmark *et al.*, 2000). Virulence gene regulators can affect the expression of target genes directly or indirectly.

Between 45 and 66 genes encoding *S. aureus* lipoproteins were predicted out of almost 2,500 open reading frames of *S. aureus* genome (Babu *et al.*, 2006, Sibbald *et al.*, 2006), but most of the lipoprotein remained as predicted lipoprotein or with unknown functions (Schmaler *et al.*, 2010). One of the problems in determining gene functions in pathogenesis is that inactivation of limited genes does not completely remove the *S. aureus* pathogenicity, suggesting that virulence is a multi-factorial process (Said-Salim *et al.*, 2003).

1.3.1 *Staphylococcus aureus* global regulation genes

Virulence genes expression in general tends to be influenced by a different factors including, the concentration of autoinducing peptides and by bacterial density, pH and CO_2, and each of these signals controls different regulatory systems. The ability of *S. aureus* to produce a number of virulence factors including those related to life threatening infections, antibiotic resistance and survival in distinct adverse environments and antibiotic resistance genes is controlled either by two component systems (e.g., *agr*, *saeRS*, *srrAB*, *arlRS*, *lytSR*) and/or by transcriptional regulators (e.g., *sarA*, *sigB*, *sar* family genes, *tcaRA*) (Ballal *et al.*, 2009, Gordon *et al.*, 2013).

1.3.2 *Agr* (accessory gene regulator)

A huge effort has been directed toward identifying genes and regulatory mechanisms associated with *S. aureus* virulence factors as previously explained in section 1.1.5. Expression of virulence genes is highly coordinated and is generally controlled by global regulatory elements (Chien *et al.*, 1999). The dual action of *agr* global regulator, which acts during the post-exponential phase, has been shown to inhibit the transcription of some cell wall-associated proteins e.g. coagulase, protein A and fibronectin binding protein (Tseng *et al.*, 2004). In contrast, *agr* activates many exoproteins such as beta-hemolysin, alpha-toxin, TSST-1 and leucotoxins (Dinges *et al.*, 2000). *Agr* mutation identified 104 genes including 20 putative virulence genes that were up-regulated and 34 genes were down-regulated by agr, in the same experiments a *sarA* mutation altered expression of

~120 genes, to show 76 genes up-regulated and 44 were down-regulated (Dunman *et al.*, 2001). The effector of *agr* at the transcriptional and translational levels is a RNA molecule, RNAIII-*agr* intergenic region, which modulates virulence factor expression (Novick *et al.*, 1993).

The *agr* locus consists of 5 genes, *agrA*, *agrC*, *agrD*, *agrB* and *hld*, however its consist of two divergent transcripts RNAII and RNAIII with sizes of 3 kb and 0.8 kb, respectively, which are under the control of two main promoters P2 and P3 (Peng *et al.*, 1988). RNAII transcript driven by P2 promoter and encodes four gene operons, agrBDCA, RNAIII effective molecule is responsible for up-regulation of extracellular protein synthesis and down-regulation of cell wall associated protein production during post exponential phase (Chien *et al.*, 1999).

The consistent regulation of extracellular and cell wall virulence factors during growth has hinted at the contribution of global regulatory system in *S. aureus*, e.g. in the exponential phase, cell-wall adhesive functions proteins are actively synthesized to correspond with the tissue binding and colonization stage of infection, these proteins known as MSCRAMMs (microbial surface components recognizing adhesive matrix molecules), while post-exponential phase, the expression of cell wall proteins is decreased and the synthesis of extracellular toxins and enzymes is increased (Cheung *et al.*, 2004). Both *agr* and *sarA* global regulators have been shown to co-ordinately regulate the transition from exponential to post exponential bacterial protein expression (Novick, 2003).

Some virulence gene can be under the influence of several regulators or regulatory systems, which influence each other to ensure expression of the target gene within appropriate conditions. Expression of these factors is controlled by multiple regulatory systems such as the accessory gene regulator (*agr*) and staphylococcal accessory regulator (*sarA*) loci, *S. aureus* accessory regulatory protein (*SarA*) influencing both exoprotein and cell surface protein expression (Cheung *et al.*, 1992).

1.3.3 Sar protein family structure

Sar locus encoded within a 1.2-kb DNA fragment, encompasses three overlapping transcripts (*sarA*, *sarB* and *sarC*) each of which terminates at the same site and encodes for the SarA protein, these transcripts, designated *sarA* (0.58 kb), *sarB* (0.8 kb) and *sarC* (1.2 kb), have similar 3’ ends but within three distinct promoters (Manna *et al.*, 1998). SarA

has been shown to promote production of many extracellular and cell wall-associated proteins, while inhibiting the transcription of protein A and protease genes, also it is required for full agr expression in *S. aureus* (Cheung *et al.*, 1997). SarA in *S. aureus* is constitutively produced during cell growth phase, meanwhile expression of each *Sar* transcript occurs in a growth phase dependent manner; the highest expression level of *sarA* and *sarB* are mainly transcribed during the log phase, in contrast, *sarC* is predominantly expressed during stationary phase of growth (Bayer *et al.*, 1996). SarA protein regulate some target genes by directly binding to gene promoters or indirectly by downstream effects on regulons e.g. binding to *agr* promoter or by stabilising mRNA during exponential phase (Roberts *et al.*, 2006) . *SarA* is the major regulatory molecule of this locus and mediates its elect both directly by binding to target gene promoters (e.g. *agr*, *hla* and *spa*) and indirectly via the downstream elect on other regulons that *sarA* locus controls the expression of over 100 target genes (Dunman *et al.*, 2001).

Locus *sarA* encodes a 372 bp open reading frame with three upstream promoters (P2, P3, and P1) driving three overlapping transcripts, each of them coding for the 14.5 kDa SarA protein (Cheung. *et al.*, 1997). The 372 bp ORF *sarA* together with extensive 800 bp upstream sequence is present within the main transcript *sarB* (Chien and Cheung, 1998). The crystal structure of a *SarA* DNA complex has been explained and indicates that *SarA* mediates DNA supercoiling form (Roberts *et al.*, 2006). *SarA* similar to *SarR* structures, it is diametric winged helix character including each monomer consisting of 5 α-helices, 3 β-strands and several loops (α1α2-β1α3α4-β2β3-α5), *SarA* dimer possesses a central helical core and two winged helix motifs, each winged helix motif has a helix-turn-helix motif (α3α4) and a β-hairpin turn wing (β2β3), both of them employed as putative DNA binding domains (Schumacher *et al.*, 2001).

1.3.4 Staphylococcal mechanisms and genes encoding lipoproteins for iron and manganese uptake

Uptake of nutrients from the environment by bacteria is an essential process that supplies bacteria with different elements and allows them to recognise the environmental conditions. Determining the complete genome sequences for some strains has allowed detection of those genes encoding a number of iron transporters, searching within genome database using amino acid sequences of known iron transporters. Iron depletion conditions in host body fluids serve as a most important environmental signal to bacteria to express

virulence factors (Trivier and Courcol, 1996). Iron uptake system network usually based on a member of the ATP-binding cassette (ABC) family of transporters for importing iron or complexed-iron across the membrane, this multi subunit complex structure contains; (i) integral membrane proteins, which act as a permease; (ii) peripheral membrane ATP binding proteins that hydrolyse ATP; and (iii) extracellular substrate binding proteins (SBPs) that work as receptors (Williams *et al.*, 2004).

Structure of SBP systems is slightly different between Gram-negative and Gram positive bacteria, in Gram-negative bacteria, SBPs are secreted into the periplasm and are retained in this space by the outer membrane, but in Gram-positive bacteria, with the absence of outer membrane, are required to tether these proteins to the plasma membrane via a lipid anchor or by incorporating with a specific integral membrane element of the transporter (Tam and Saier, 1993). Two different mechanisms of iron uptake in *Staphylococcus* spp have been described, siderophore or non-siderophore based pathways to use a range of available iron substrates including a release of low molecular-weight high-affinity iron scavenging molecules known as siderophores, which can bind to specific receptors if combined with iron molecules including staphyloferrin A and B, and in addition *S. aureus* produces a siderophore called aurochelin (Brown and Holden, 2002).

Different studies have identified nine iron regulated lipoproteins (IRLPs) from approx. 55-70 total lipoproteins expressed by *S. aureus*, SirA, HtsA, SstD, FhuD1, FhuD2, IsdE, FepA, SitC and Opp1A, and their important functions in bacterial pathogenesis and immune stimulation (Sheldon and Heinrichs, 2012, Wright and Nair, 2012). Ferric uptake regulator (*Fur*) gene is a homodimeric metalloprotein which acts as a transcriptional regulator of iron homeostasis in various bacteria (Andrews *et al.*, 2003). Other IRLPs staphylococcal iron transporter, SitABC/MntABC proved to play a role in iron uptake and homeostasis (Sheldon and Heinrichs, 2012).

Well distinguished system of iron uptake in *S. aureus* is the ferric hydroxamate uptake (Fhu) system, encoded by up to 5 genes including *fhuC* (ATPase), *fhuB* and *fhuG* (to assemble membrane embedded permease) located in an operon that encodes an ATP-binding cassette (ABC) transporter, FhuC gene is a predicted ATPase, while *FhuB* and *FhuG* (high-affinity receptors) are membrane spanning proteins (Sebulsky *et al.*, 2000). Lipoprotein FhuD2 (ferric hydroxamate uptake) was first discovered as a ferric hydroxamate siderophore binding protein that involved iron uptake via a dedicated ATP

binding cassette (ABC) transporter (Sebulsky and Heinrichs, 2001). *S. aureus* RN6390 strain carries at least 5 iron regulatory genes, their products regulating the ferric hydroxamate uptake mechanism, three gene operon fhuCBG encodes the ATPase, however *fhuD2* codes lipoprotein *fhuD* are the fifth gene involved in this transport scheme (Sebulsky *et al.*, 2004).

S. aureus can also consume a range of siderophores produced by other bacteria such as enterobactin, a catechol type siderophore and some hydroxamate siderophores e.g. ferrichrome and aerobactin (Sebulsky and Heinrichs, 2001, Sebulsky *et al.*, 2004). Predominant lipoprotein SitC one of staphylococcal iron transporters; SitABC (Cockayne *et al.*, 1998), PrsA, peptidyl-prolyl cis-trans-isomerase a part of protein folding system and OppA, oligopeptide permease (Stoll *et al.*, 2005).

Manganese (Mn) uptake has been shown to be essential for the growth and survival of many living organisms and play specific cellular roles *in vivo* pathogenicity as a cofactor in enzymes for metabolism, catabolism and photosynthesis (Jakubovics and Jenkinson, 2001). In *S. aureus* Mn has non-enzymatic role in the protection of the cell from oxidative stress factors (Clements *et al.*, 1999). Control of iron, zinc and Mn uptake are mediated by members of Fur and DtxR protein families (Hantke, 2001). ABC permeases transport binding protein in Gram-positive bacteria has lipoprotein modification (Sutcliffe and Russell, 1995). The MntABC operon is composed of three genes, MntA the ATP-binding protein, MntB the integral membrane transporter and MntC metal binding lipoprotein (Nielsen *et al.*, 2011). The MntC lipoprotein a part of the MntABC operon is involved in the uptake of manganese and/or zinc in *Staphylococcus* spp for cellular processes and host-bacteria interactions (Claverys, 2001). Horsburgh, *et al*. showed that MntABC operon mutant *S. aureus* had lowered intracellular manganese concentration (Horsburgh *et al.*, 2002). *MntC* gene was highly expressed early during *S. aureus* infection stages of murine bacteraemia model, active immunization with *MntC* mutant strain was shown a significant reduction of *S. aureus* acute bacteraemia *in vivo* (Anderson *et al.*, 2012).

1.4 A Database of Bacterial Lipoproteins (DOLOP)

A Database with a various features based on bioinformatics analysis tools are available to predict bacterial lipoproteins using predicted features of an N-terminus lipobox and lipoprotein signal peptides and comparison of different predictive algorithms has been

developed (Babu *et al.*, 2006). Each organism has a set of different numbers of lipoproteins and to identify their various bioinformatics features in a large data set a proper algorithm are needed to predict the lipoproteins based on their amino acid compositions by using available updated database of sequenced genomes with a higher degree of confidence. The best predictive rules on Gram-positive bacteria lipoproteins were suggested algorithm G+Lpp (Sutcliffe and Harrington, 2002), in contrast, another trained set of predictive rules suggested algorithm for Gram-negative bacteria to predict lipoproteins by looking at the signal sequence features and proposed as LipoP (Juncker *et al.*, 2009). The comparative analysis of DOLOP database for bacterial lipoproteins information, based on the data of molecular bioinformatics of different functional lipoproteins, these predicted lipoproteins from about 234 completely sequenced bacterial genomes, as a result of all these works, the numbers of bacterial lipoproteins that have been identified would exceed several thousand from a variety of bacterial sources (Babu *et al.*, 2006). Database of bacterial lipoproteins (DOLOP) is available at (http://www.mrc-lmb. cam.ac.uk/genomes/dolop). However, unfortunately this database was not kept up to date as it was last modified in April 2005, but it is provided a useful starting point to study staphylococcal lipoproteins as data from 6 strains of *S. aureus* were included.

The availability of complete genome sequences for some *S. aureus* strains can provide reliable information to discover lipoproteins sequences, depending on the predictive basics. Also, provide some evidence about possible functions by identifying protein domains predicted lipoproteins. Furthermore, by using protein sequences database with the advantage of a combination of keywords, limits research results to only one species (Bairoch and Apweiler, 2000). Some *S. aureus* lipoproteins were not detected by DOLOP precomputed features-tools because the regular expression pattern allowing detection of lipidation sequence ([LVI] [ASTVI] [GAS] [C] lipobox) is too stringent. The LipoP method to predict lipoprotein signal peptides in Gram-positive is more permissive, however using a single tool may result in errors (Bagos *et al.*, 2008). This suggests the best approach is to combine the various features-based methods available and compare the final results. Determination of outer membrane localization machinery (LOL) for lipoproteins and the effect of amino acids in the modifiable cysteine in the mature sequence in their recognition and amino acid residues at +3 and +4 positions were found to affect the membrane localization of native lipoproteins (Terada *et al.*, 2001). Another method to assign lipoprotein predictions is the Prosite position-specific matrix PS51257, this system

has been used to predict lipoproteins encoded by Gram-negative bacterial genomes (Hulo *et al.*, 2008).

1.4.1 Proteome analysis of *S. aureus*

Proteomics allows a better understanding of biological and pharmacological structure of bacterial lipoproteins. proteomics provide a perfect complement to comparative genomics analyses, by giving a large view of protein profiling and their ability to visualise changes in protein expression in *S. aureus* (Gatlin *et al.*, 2006, Kohler *et al.*, 2005). Classical proteomics techniques have shown changes of the metabolism and pathogenic by visualisation of main metabolic pathways in a simple proteomics procedures. Proteomes are characterized in the past by two ways (i) protein microarrays and (ii) protein staining of two-dimensional gel electrophoresis (2DE) which resolves the complex protein sample in two dimensions based on molecular weight and pI (isoelectric point). Both methods were has variable reproducibility and low characterization of the less abundant proteins (Janini and Veenstra, 2002, MacBeath, 2002).

It is a highly recommended technique for separation and quantitation of lipoproteins, and has much to contribute to the experimental analysis of bacterial cell proteins (Brady *et al.*, 2006). However, after gel-free and mass spectrometry techniques have been introduced to proteomics experiments, 2-DE has become the second choice in proteomics approach. Gel-based techniques have some limitations including a poor reproducibility, limited dynamic range, sensitivity, and its time consuming and there are many technical problems that can occur (Janini and Veenstra, 2002).

A different proteomics strategy coupled gel electrophoresis with mass spectrometry, proteins spots resulting from gel separated by 2-DE, subsequently analysed by mass spectrometry machines (MALDI-TOF-MS or by LC-MS/MS), nevertheless, the drawback of protein analysis by mass spectrometry is the limited molecular information for the intact secreted proteins (Ravipaty and Reilly, 2010). In *S. aureus* 2-DE/MS proteome mapping has been used successfully to investigate hundreds of proteins in both, intracellular (cytoplasmic) (Kohler *et al.*, 2005) and extracellular (secreted) proteins (Kohler *et al.*, 2003). The physico-chemical properties of proteins and analytical window of 2-D gels can affect the abundance of individual proteins on the gel, in quantitative proteomics study on

S. aureus COL, 683 proteins were found to be different between proteome patterns of analytical windows of pI 4-7 and in the pI range of 6-11 (Becher *et al.*, 2009b).

The introduction of mass spectrometry (MS) techniques with high resolution and mass accuracy, and using gel-free sample preparation methods, has achieved significant analytical improvements. Gel-free MS-based proteomics approaches need smaller samples, in 2D gel-based proteomics μg quantities of sample proteins are needed, but with modern gel-free proteomic approaches ng of samples can be used (Schmidt *et al.*, 2010). Within a short time, the new approaches of LC/MS allow identifying a hundreds of peptides within a single run providing significantly more information (Thakur *et al.*, 2011).

The three major components of mass spectrometry instruments are an ionisation device, a mass analyser, and a detector, all ions generated by the ionisation source are separated by the analyser that separates them according to their mass-to-charge (M/Z) ratios, then the selected ions passed towards the detector where they are finally registered (Ravipaty and Reilly, 2010). The whole genome sequencing of important microbial pathogens unlocked the codes to identify different proteins via mass spectrometry as theoretical masses can be calculated and found in an available database. To analyse proteins by mass spectrometry, the size and polarity are impediment for some of them, but with some site specific proteases can overcome this issue by cutting proteins into small peptides (Mann *et al.*, 2001).

In *S. aureus* relatively quantitative combination strategy has been used effectively for proteomics identification and has characterised the features of extracellular proteins, in order to find out the regulatory network involved in pathogenicity (Ziebandt *et al.*, 2004). Meanwhile, Cordwell, *et al*. mapped the cytoplasmic proteins of two strains (COL and 8325) and covered about 12% of *S. aureus* proteome (Cordwell *et al.*, 2002). Most of the hydrophobic proteins remain undetected, either because they precipitated during isoelectric focusing, or due to their low abundance (Nandakumar *et al.*, 2005). Transcriptional analyses of *S. aureus* showed the main regulation process of virulence factors expression under different environmental conditions but the proteome during host infection are variant due to protein production can be regulated post translationally (Cordwell *et al.*, 2002). Most of proteomics studies on *S. aureus* virulence in infected tissues have been performed by applying culture stress conditions that challenges that the microbe encounters during infection, these studies do not show the pathogen activities during infection challenges

and limitations imposed by host interference (Bumann, 2010). Proteomics revolution in biotechnology has led to discovering of many important proteins involved in different approaches of bacterial infection and diseases, this advance technology allowed the scientists to find out wide proteome for a number of pathogenic organisms.

1.5 The Nematode *Caenorhabditis elegans*

Caenorhabditis elegans (*Caeno*, recent; *rhabditis*, rod; *elegans*, nice), is a free-living, soil-dwelling and self-fertilizing non-parasitic nematode, microbivore where it feeds on a wide range of bacteria and fungi, microbes passed into the pharynx and crushed by a cuticular (grinder), and pumped through to the intestine (Nicholas and Hodgkin, 2004b), some of these microorganisms are identified pathogens, capable of infecting and even fatal to *C. elegans*. From a practical standpoint, it is easily maintained and manipulated in basic laboratories, can be propagated on agar plates or in liquid media and can live up to 3 weeks at room temperature. *C. elegans* transparent body facilitate visualizing and dissecting internal features using simple dissecting microscopes. Their small in size (1.5 mm long), short generation time (produce 300 genetically identical progeny in a 3 days life cycle at 25°C), ease of handling, powerful genetics, invariant developmental lineage, small and completely sequenced genome (100 Mb) give it an abundant advantage to use as valuable effective model host of research (Powell and Ausubel, 2008).

Biological data and entire genomic sequence for *C. elegans* are assembled in freely accessible public databases and numerous strains including mutant strains are available through scientific resources, also its small size, ease of handling, short generation time and powerful genetics give it great advantage to be employed as host model (Schulenburg *et al.*, 2004). Their use does not raise ethical issue related to use of animals as biological models. Various features of interaction of *C. elegans* with pathogens have been described in immunogenetic studies, also many model host organisms provide a number of useful experimental advantages for discovering genes and genetic pathways involved in biological processes (Millet and Ewbank, 2004, Nicholas and Hodgkin, 2004a). Transgenic *C. elegans* strains were easily created using microinjection DNA and worms transparency renders the use of fluorescent reporter genes *in vivo* (Aballay *et al.*, 2000).

1.5.1 Life cycle of *C. elegans*

Well-fed *C. elegans* can develop and reproduce at a wide range of temperatures, however development stops at temperature below 8°C and the worms become sterile above at 27°C, it has a life cycle approx. 3 d under ideal conditions and their average life span is 3 weeks at 20°C (Félix and Braendle, 2010). *C. elegans* life cycle has six stages: consists of embryonic stage, four postembryonic larval stages and a reproductive mature stage with optional reversible, developmental arrest at some larval stages, the life cycle of *C. elegans* shown in figure 6.

Hermaphrodites, *C. elegans* produce both oocytes and sperm and reproduce in the absence of males by self-fertilization. In favourable conditions, males arise infrequently (0.1%) by spontaneous non-disjunction in the hermaphrodite germ line and at higher frequency (up to 50%) through mating (Hodgkin, 1986). Hermaphrodite worm produces larger number of oocytes and approximately 300 sperm, to generate about 300 progeny by self-fertilization in 4-5 days and more than 1000 progeny by a male fertilization (Ward and Lamunyon, 1995). Embryonic stage divided into two phases, the first phase requires cells proliferation and organogenesis, however elongation, morphogenesis and cuticle synthesize occur in the second phase, finally the animal hatches into an L1 larva (Edgar *et al.*, 1994).

When environmental conditions are favourable L1 larvae develops through 4 larval stages (L1 to L4) to adult within 3 to 5 d when conditions are favourable e.g. food, temperature and crowdedness. During post-embryonic maturation stage the structures of sexual reproduction (gonad, vulva and sex muscles) also the neurons and epidermal tissues are completed in this stage. In unfavourable conditions *C. elegans* larvae are capable of interrupting their reproductive growth by arresting at an alternative L3 stage termed dauer which is characterized by developmental quiescence, stress resistance and more adapted for long-term survival (Wang and Kim, 2003). Sexual immature, starved and non-aging dauer larvae can survive for up to three months in unfavourable conditions, then resume development when conditions and environmental factors changed (Wood, 1988). Availability of food after hatching triggers post-embryonic development and cell divisions resume, post-embryonic growth begins after 3 h of hatching, but in the absence of nutrition L1 larvae arrest development until food becomes available (Sulston and Horvitz, 1977).

C. elegans has two sexes, hermaphrodites and males, hermaphrodites can self-fertilize or mate with males but cannot fertilize each other (Fischer, 1988). *C. elegans* genome has five pairs of autosomes chromosomes and one pair of X chromosome, worm sex determination is defined by the copy number of X chromosome; hermaphrodites have two copies (XX) and whereas males have one (XO) (Riddle *et al.*, 1997).

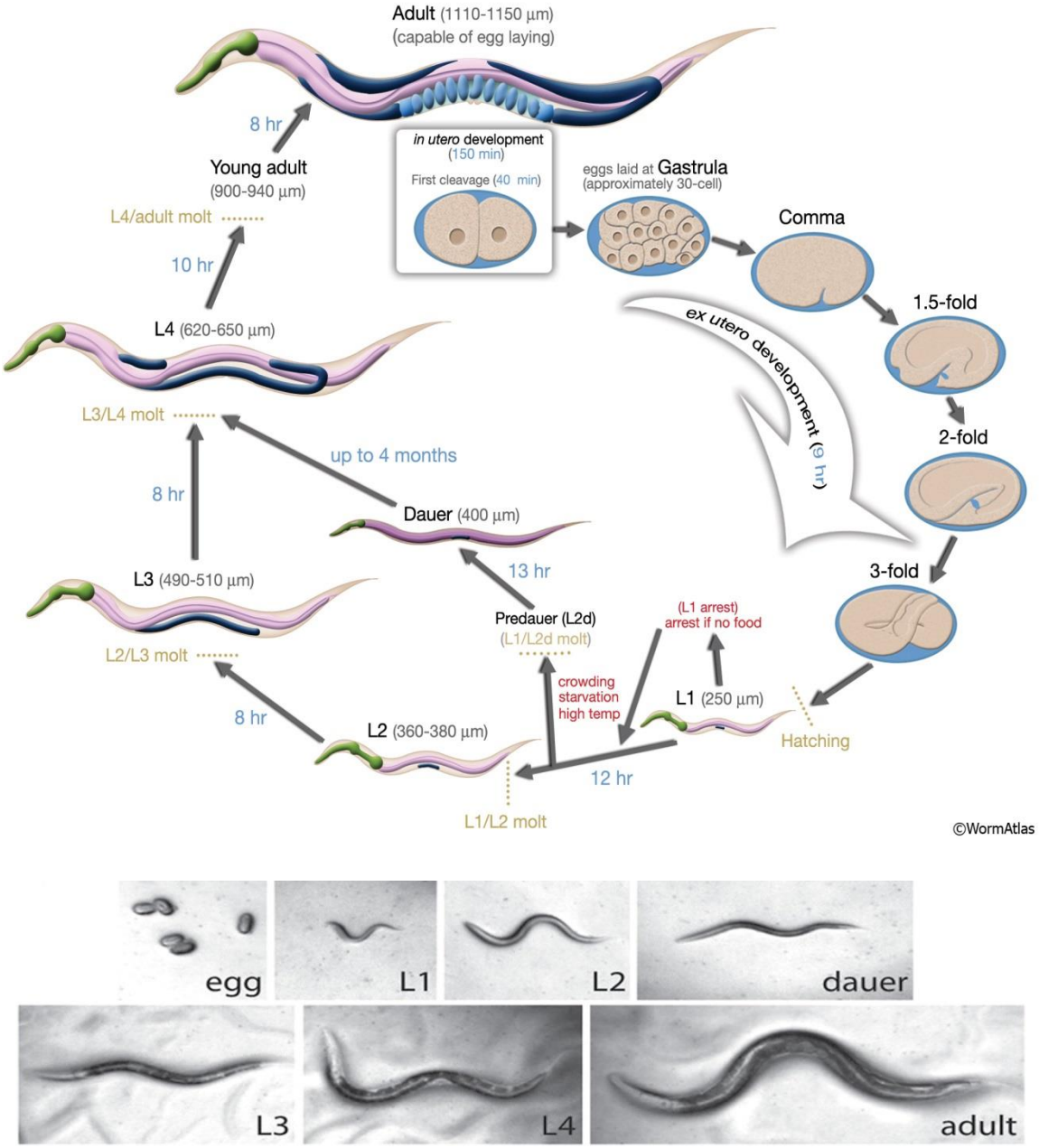

Figure 6. Life cycle of *C. elegans* at 22°C

Starting with fertilization time followed by each name and length of time that animal spends at different stages. Worm Atlas, Altun. (http://www.wormatlas.org/ver1/handbook/anatomyintro/anatomyintro.htm).

As *C. elegans* feeds on bacterial lawns, this attribute the model host in two ways; it provides a convenient route for infection by microbial pathogens, and bacterial feeding can be used for the application of RNA interference (RNAi) based gene silencing, by using *E. coli* strains expressing dsRNA (Kamath and Ahringer, 2003).

C. elegans is exposed to various bacterial and fungal pathogens in their surrounding environment and expected to have evolved effective defense mechanisms to face this infection. *C. elegans* is the most attractive model host in many biological researches, especially in developmental biology and neurobiology, these researches rely on the fact that virulence factors relevant for infection of mammals are similar for full pathogenicity during infection of nematodes (Schulenburg *et al.*, 2004). It is well suited parasite for mutagenesis and forward genetic analysis and has a fully sequenced genome (Strange, 2003). *C. elegans* is a simple nematode model host for studying the relationship between the animal innate immune system and many bacterial and fungal pathogenesis (Powell and Ausubel, 2008).

When *C. elegans* feeds on the standard laboratory diet *E. coli* OP50 strain, the pharyngeal grinder crushes all food and no live bacteria were found in the worm intestine (Aballay *et al.*, 2000, Garsin *et al.*, 2001), however, if the pathogenic microbes pass intact through the pharynx and accumulate in the intestine, this can cause infection. A large number of different microbial pathogens have been revealed to infect *C. elegans*, including Gram-positive bacterial pathogens *Staphylococcus aureus* (Sifri *et al.*, 2003), *Microbacterium nematophilum* (Hodgkin *et al.*, 2000) and *Enterococcus faecalis* (Garsin *et al.*, 2001); Gram negative bacterial pathogens *Pseudomonas aeruginosa* which was the first bacterium shown to be able to infect and kill *C. elegans*; (Tan *et al.*, 1999), *Salmonella enterica, Salmonella typhimurium* (Aballay *et al.*, 2000), *Serratia marcescens* (Kurz and Ewbank, 2000), and *Yersinia pestis* (Darby *et al.*, 2002); and the pathogenic yeast *Cryptococcus neoformans* (Mylonakis *et al.*, 2002), moreover, some putative fungal pathogens to *C. elegans* have been recognised, including *Drechmeria coniospora* (Jansson *et al.*, 1985). *C. elegans* has shown many ways to survive from the harmful effects of these infections by sensing and distinguishing between bacterial compounds in their environments (Pradel *et al.*, 2007).

Infected *C. elegans* show visible signs of illness e.g. bacterial biofilm formation, tail swelling and massive internal accumulation of microbes or die more quickly with mean

lifespan between 1.5-7 days. In contrast, with non-pathogenic *E. coli* OP50 the mean lifespan time is 15-17 days (Powell and Ausubel, 2008). *S. aureus* was amongst the first models used in *C. elegans* immunity research experiments and *S. aureus* A003 strain was shown to exhibit a high level of nematocidal activity with average time for half of the worms to die of ~ 2 days (Garsin *et al.*, 2001). A significant killing activity of 23 clinical *S. aureus* isolates were examined and revealed to kill more than 70% of nematodes during the course of ~ 5 days, also all *C. elegans* larval stages were killed by *S. aureus*. The main molecular mechanisms of pathogenicity are not fully understood but accumulation of bacteria in infected nematodes digestive tract was observed (Sifri *et al.*, 2003). Worms fed on *S .aureus* appeared normal for the first 16 to 20 h including nematode locomotion detected visually, pharyngeal pumping and foraging, however after 24 to 48 h all of these activities gradually decreased until the worms became immobile and died (Sifri *et al.*, 2003). Adult worms proved to be more susceptible to *S. aureus* infection than all larval stages (Bae *et al.*, 2004). Various bacterial virulence genes required for the nematode were also important in other model hosts (Tan *et al.*, 1999).

The dead worms lost all cellular appearance architecture and appeared as ghosts in bacterial lawn, furthermore, "bag of worms" phenomena were observed in the infected death nematodes, when eggs of a gravid hermaphrodite hatched internally and the resulting L1 larvae consumed their parents. This matricide might be occur because infected nematodes become too weak to lay eggs normally (Garsin *et al.*, 2001, O'quinn *et al.*, 2001). Examination of infected worms under Nomarski differential interference contrast microscopy showed clear distention of the intestinal lumen with visible intact bacteria, although worms fed *E. coli* OP50 had slender intestinal lumen with no visible intact bacteria (Sifri *et al.*, 2003). *P. aeruginosa* PA14 strain expressing *Aequorea victoria* GFP was able to kill worms after 48 h of infection, clear green fluorescence pigments was observed within the lumen of worm intestines, indicating the accumulation of bacteria (Tan *et al.*, 1999).

1.5.2 Infection routes of *C. elegans*

Microbial pathogens can harm *C. elegans* either by colonizing the intestine or adhere to nematode cuticle and decrease its lifespan, while other pathogens produce toxins that kill *C. elegans* without any directly contact with the worms. The main two routes of infection of *C. elegans* through pharynx or the epidermis. Some organisms cause the pathogenic

effect after ingestion by worms, often exerting their effect in the anterior part of the intestine where they may establish an intestinal infection (Gravato-Nobre and Hodgkin, 2005). Intact microbes must manage to pass through upper digestive tracts and surviving any disruption by pharyngeal grinder action, start to proliferate within the gut, causing distension and damaging the intestinal epithelium cells, leading to mortality associated with epithelial disruption (Köthe *et al.*, 2003, Mylonakis *et al.*, 2002), Gram-negative bacteria *Salmonella typhimurium* have been shown to establish intracellular infection in the intestinal cells (Jia *et al.*, 2009), other important pathogens was found to have significantly detrimental activity by their production of bacterial toxins, this was reported with *P. aeruginosa* PAO1 (Gallagher and Manoil, 2001) and *Bacillus thuringiensis* 010 strain (Zhang *et al.*, 2014).

It is relatively simple to substitute non-pathogenic food source of *C. elegans* with pathogenic bacterial strains, this allow analysing the involved virulence mechanisms and host defences. *C. elegans* are able to distinguish metabolites, autoinducers, odours and pathogenic bacteria (Beale *et al.*, 2006). Due to chemical and mechanical features of *C. elegans* skin, only limited bacteria and fungi were proved to bind to or degrade the extracellular exoskeleton of worm. A few pathogenic microbes have been shown to adhere to and then penetrate the cuticle and grow into the epidermis, for example the fungus *Drechmeria coniospora* were capable of adhering and causing mouth and vulval infections (Jansson, 1994). The bacteria *Yersinia pestis* and *Yersinia pseudotuberculosis* formed biofilms on the heads of the nematodes, the dense mass of bacteria prevented feeding (Joshua *et al.*, 2003).

According to JebaMercy and Balamurugan, virulence of pathogen can be influenced by preinfection of *C. elegans* with other microbes, this immune challenge was shown when preinfected *C. elegans* with pathogenic *S. aureus* increased the vulnerability of host by significantly reducing the life span and increasing the chance for opportunistic pathogen *Proteus mirabilis* to be more pathogenic (JebaMercy and Balamurugan, 2012).

1.5.3 *C. elegans* immunity system

Many conserved innate immune defence have been revealed by investigate host/microbes interactions using vertebrate and invertebrate models. The discovery that the innate defence mechanisms of invertebrates, vertebrates and some plants shared similarities was a major development in immunological studies. All vertebrates have employed complex immunity system used to repel against pathogenic microbes, this complex immune mediated by two highly co-dependent immune systems known as innate and adaptive, whilst, adaptive immune system seems to be absent from all invertebrates and consequently its rely on their innate immune mechanisms in their response against infections (Aballay, 2013, Gravato-Nobre and Hodgkin, 2005).

Innate immune defence is the early barrier to infectious agents and acts immediately, innate immune responses are generally initiated via host recognition the conserved structures of pathogens as foreign body known as pathogen associated molecular patterns (PAMPs) e.g. lipopolysaccharide, peptidoglycan or modified nucleic acids, PAMPs sometimes known as microbe associated molecular patterns (MAMPs), moreover, recognition of damage-associated molecular patterns (DAMPs), for example ATP and monsodium urate crystals can enhance activation by PAMPs, but also may be enough to initiate the innate immune activities (Twumasi-Boateng and Shapira, 2012). Different classes of pattern recognition receptors (PRRs) responsible for recognizing PAMPs or DAMPs molecules including many proteins such as NOD-like receptors (NLRs), Toll/Toll-like receptors (TLRs) and RIG-I-like nucleotide recognition receptors (RLRs) leading to the activation of intracellular signalling pathways (Kumar *et al.*, 2011). These pathways can mediate upregulation of expression of specific defensive genes, the effector molecules product serve to limit infection or destroy invading pathogens.

Non-vertebrate, *Drosophila melanogaster* (fruit fly) was an important host model to discover Toll family receptors the most widely distributed molecules involved in mammalian defence, non-vertebrate host model *D. melanogaster* become a model to study host/ microbe interactions (Lemaitre *et al.*, 1996). NLRs take part in antimicrobial action in both animals and plants (Ting *et al.*, 2008). In vertebrates, TLRs prominent class of PRRs, can sense the outer membrane components of bacteria (Akira *et al.*, 2006). *C. elegans* avoid pathogenic elements through a process mediated by a Toll receptor (Pujol *et*

al., 2001). Triggering of long lasting adaptive immunity which has immunological memory against infectious agents is usually initiated via the innate immune system.

1.5.4 *C. elegans* defence strategies

C. elegans has three important tactics of defences against microbial pathogens, the first strategy is the avoidance behaviour against certain pathogens, worms are able to differentiate between different bacterial lawns and then they can evade possibly pathogenic strains (Andrew and Nicholas, 1976). Olfactory sense, G protein coupled receptors and Toll‐like receptor (TOL‐1) were involved in triggering the avoidance behaviour against *Serratia marcescens* pathogen (Pradel *et al.*, 2007). Many soil bacteria have employed the defense strategy of releasing toxic materials, e.g. *Pseudomonas fluorescens* produces extracellular secondary metabolites to act as a defense mechanism against bacterivorous nematodes (Neidig *et al.*, 2011).

The second line of protection against any microbial attacks is a strong cuticle which constitutes the exoskeleton of worm, when a pathogen cannot be avoided. The third mechanisms of *C. elegans* defence is the complex inducible defence system including many signalling cascades that regulate the production of antimicrobial peptides and proteins in a pathogen and tissue specific way (Engelmann and Pujol, 2010). Mechanisms of pathogenesis and infection among different microbes models would be also carefully observe the differences in the immune response that been used in *C. elegans*, these differences should relative to the disease process as it is associated with toxins and/ or infection (e.g. *P. aeruginosa* and *S. aureus*) or by establishing an infection (e.g. *M. nematophilum*) (Schulenburg *et al.*, 2004). Osanai, *et al*. demonstrated that TSST-1 and SEC toxins produced by *S. aureus* were playing a role in *C. elegans* avoidance which is mediated by 5-HT signalling pathway (Osanai *et al.*, 2012). Avoidance of *Pseudomonas aeruginosa* by *C. elegans* was subsequent to activation of TIR-1-NSY-1-SEK-1 signalling pathway in ADF neurons which neighbour pharyngeal pumps and are responsible for the avoidance behaviour of *C. elegans* (Shivers *et al.*, 2009). The scaffold protein Toll/interleukin-1 receptor (TIR-1) was also important for identify pathogenic microbes and pathogen-associated molecular patterns (Couillault *et al.*, 2004).

1.5.5 The specific requirements for pathogenesis

There are significant differences in the genetic requirements of *S. aureus* to infect nematodes and humans, one of these significant difference is the temperature during infection, as the worms optimum growth temperature is 25°C, however temperature of mammals before infection is 37°C and during the later stages of infection is between 32°C and 43°C, these temperature changes indicating changes in genes expression and their virulence affects (Vlach *et al.*, 2000). Another study proved that some genes required for pathogenesis of one staphylococcal infection were unnecessary in a different disease model (Schwan *et al.*, 1998). Lack of an adaptive immune system in *C. elegans* is another drawback to use it as a model host, also *C. elegans* seems to have less or no specialized cells similar to scavenging phagocytic neutrophils and macrophages (Gravato-Nobre and Hodgkin, 2005).

1.5.6 *C. elegans* antimicrobial peptides and proteins

Generally *C. elegans* produces a set of antimicrobial peptides and proteins to defend itself against microbes. *C. elegans* has a numerous candidates of antimicrobial effector molecules, some of them were equivalent to known antimicrobial factors in nematode species while others were transcriptionally regulated upon infection models, which seem to be produced rapidly and do not have similar structures even in other nematodes (Ewbank and Zugasti, 2011).

1.5.6.1 Caenopores

Saposin like proteins (SPPs) also known as caenopores, is a large family, with 28 coding genes constitutes at least 33 different *C. elegans* antimicrobial proteins, caenopores structure belong to SAPLIP (saposin like protein) superfamily as exemplified by the structure of caenopore-5 (SPP-5) (Aylin *et al.*, 2012). Caenopores have a similar structure and function characteristics with a range of organisms from amoebae to mammals such as the pore-forming cytotoxic proteins that can kill bacteria (Roeder *et al.*, 2010). These proteins are generally expressed in *C. elegans* intestine and their expression can be triggered by specific pathogenic bacteria but their mode of action has not been fully studied (Ewbank and Zugasti, 2011).

1.5.6.2 Lysozymes

Lysozymes are another class of molecules show a significant role in both vertebrate and invertebrate immunity. Lysozymes are a group of important putative antimicrobial proteins in *C. elegans*, their expressions in the intestine was reported to be induced by specific infections (Alper *et al.*, 2007). The function of enzyme is attacking the peptidoglycans layer of bacterial cell walls mainly the Gram-positive bacteria. The immune system of *C. elegans* has a set of lysozyme genes playing an important role against the invading pathogen (Shapira *et al.*, 2006). *C. elegans* lack C‑type lysozymes, but possesses 15 lysozyme genes divided into 3 classes, two of them related to protist lysozymes and one specific to invertebrates (Schulenburg and Boehnisch, 2008). Some of these lysozyme activities have been observed to be important in *C. elegans* resistance to many Gram positive and Gram negative bacteria (O'Rourke *et al.*, 2006). Lys‑7 lysozyme was induced via bacterial challenge and their inactivation has render worms to be more susceptible against *M. nematophilum* and *P. aeruginosa* (Mallo *et al.*, 2002, O'Rourke *et al.*, 2006). The exact immune defence functions of all these lysozymes remains to be determined.

1.5.6.3 Lectins

The large family of lectins also involved in innate defence mechanism in many species, *C. elegans* show a complex patterns of genes regulation of lectin, expression of some lec and clec genes was up-regulated by several pathogens, however their activation relatively to defence against different pathogenic infections (Mallo *et al.*, 2002, O'Rourke *et al.*, 2006, Wong *et al.*, 2007). Lectins were involved in microbe recognition via binding to pathogen surface sugars, others were act as opsonizing factors and/or possess antimicrobial activity, whilst others seem to play a role in masking host epitopes the targets for some specific microbial effectors (Schulenburg *et al.*, 2008).

1.5.6.4 Reactive Oxygen Species

C. elegans also has the ability to release bactericidal Reactive Oxygen Species (ROS) in response in host defence to attacking pathogens e.g. the Gram-positive bacteria *Enterococcus faecalis* which induce ROS production as an antimicrobial defense by the action of the dual oxidase BLI-3 (Chávez *et al.*, 2009). The antimicrobial activities of ROS is not specific in their action to attack the invaded microbe, this may lead to accumulation of ROS and damage some of host tissue, subsequently increasing the chance of triggering a

protective stress response in worms (Chávez *et al.*, 2007). Another evidence is that *C. elegans* can avoid cultures and culture supernatants of pathogenic bacteria *S. aureus* by rapid recognition of secretory products including staphylococcal enterotoxin C (SEC) and toxic shock syndrome toxin 1 (TSST-1), avoidance of this molecules was dependent on Toll/interleukin-1 receptor (TIR-1) and generation of 5-hydroxytryptamine (5-HT) (Osanai *et al.*, 2012).

1.6 Whole transcriptome shotgun sequencing technology (RNA-Seq)

There are thousands of genes and their products in bacterial cells that act in a complex and coordinated system to allow microbes to survive in a different environment. Traditional molecular biology methods usually work on a limited basis, which means the general depiction of whole genes function is very limited. To achieve a useful comprehensive interaction for all genes in the living organism and the expression levels of specific genes during normal growth or pathogenic processes of certain microorganism a reliable and effective tool are required to complete these aims. The advanced next-generation RNA sequencing (RNA-seq) is a modern technique that has been widely used for transcriptomic researches to provide excellent genome coverage to almost all of the expressed transcripts of an organism.

It is a process that consists in sequencing all cDNA obtained from an RNA sample to identifying and quantifying RNA molecules that are expressed at a certain time-point and under a specific set of conditions. RNA-seq was first introduced in 2008, which involves high sequencing of cDNA generated from RNA preparations to alter the general view of extent and complexity of transcriptomes and overcome some of disadvantages of existing sequencing approaches (Wang *et al.*, 2009). Next generation sequencing technology offer an enormous throughput of billions of bases per run to obtain a whole transcriptome by a single sequencing run with high coverage to randomly reveal all that can be sequenced without previous information on what has to be discovered, this high sequencing output is the key point in the sensitivity and comprehensiveness of RNA-seq analysis (Marie *et al.*, 2010). RNA-sequencing method has been used to identify and quantify molecules of non-coding short RNA (sRNA) (Vogel and Sharma, 2005). Quantitative sequencing must show relative abundance of individual RNA molecules, this abundance depends on RNA preparation and on the sequencing method, some RNA molecules tend to form secondary structures leading to less efficiently sequence which could significantly alter the

quantitative analysis (Marie *et al.*, 2010). RNA-seq can detect a different transcriptional feature including the 5′end of total RNAs (Wurtzel *et al.*, 2010).

This sensitive ultra-sequencing approach makes it possible to investigate unculturable microbes or that cannot be isolated e.g. endosymbionts (Güell *et al.*, 2011). RNA-seq has a number of advantages over microarrays, including single base pair high resolution, minimum background signal, a wide range of expression levels, enormous levels of throughput, smaller RNA material requirements and it can detect transcripts that do not match to a past sequenced genome (McClure *et al.*, 2013). Previous transcriptional analysis approaches were based on hybridization of specific oligonucleotides to particular loci for their sequence specificity by using selective primers binding to target cDNA in qRT-PCR, hybridization of cDNA to probes on microarray chips or labeled specific probes binding to RNA in Northern blotting. However, RNA-seq is various in principle through matching data to genes by sequence alignment instead, no probe sequences are specified, experimental design are similar in genome sequence and determines all transcription in an equal manner. Sequence data mapping is also more accurate than hybridization between oligonucleotides and give sequencing with high resolution, these advantages help to determine the bacterial genetic features and genetic variation between strains (Croucher and Thomson, 2010). RNA-seq quantification of genes expression is not affected by interference between the genes due to non-specific hybridization of cDNA to probes (Cloonan and Grimmond, 2008). Hybridization techniques measure genes expression levels by detection of fluorescence or radioactivity, RNA-seq has a better dynamic range for measuring variable data through revealing the amount of data matching to coding sequence, then quantified as reads per kilobase CDS length per million reads analysed (Mortazavi *et al.*, 2008). The ribosomal RNA consist the majority of total isolated RNA, depletion of rRNA molecules has been used to intensify the coverage and productivity of mRNA and non-coding RNA (Sharma *et al.*, 2010). However, quantity of RNA-seq data by abundant transcripts will need more research in some points; for example, analysing bacterial genes expression *in vivo* where host tissue RNA will be more abundant than that of the bacteria (Conway, 2003). Interestingly, it has been shown even with analysing only one RNA sample, RNA-Seq can be very useful (Meyer *et al.*, 2009). To prove highly appropriate RNA-seq datasets it should compare either technical or biological replicates to make them fit for expression studies. RNA-seq in *S. aureus* was firstly used to study small

non-coding RNAs (Beaume *et al.*, 2010), and to find the role of anti-sense transcription (Lasa *et al.*, 2011).

1.6.1 Characteristics of bacterial transcriptomes

Bacterial transcription and regulation of gene expression has been a topic of interest for many years, with different discoveries being made with the combination of modern genetics assays and highly quantitative methods. The considerable number of microbial genomes available in databanks has given the opportunity for more in depth studies of pathogenic microbes, including the post-genomics investigations and relevant biological processes. Advent of next generation sequencing technology including bacterial RNA-Sequencing has made the progress in discoveries bacterial transcriptomics and improves functional genomics experiments. RNA-Sequencing has changed the scientific view of the extent and complexity of microbial transcriptomes and provides more detailed measurement of levels for transcripts than other approaches. Bacterial transcriptomes have different characteristics than eukaryotic transcriptomes, in bacterial genome, the neighbour genes frequently overlap and consequently disturbing to differentiate between the start of individual genes transcript from the end of another (McClure *et al.*, 2013). Thus bacterial RNA-seq data analysis has different aspects of eukaryotic RNA-seq data.

Many technologies have been used successfully to infer and quantify the transcriptome, including hybridization and sequence based methods, the hybridization based technologies usually employed fluorescent labelled cDNA and custom-made microarrays or commercial high-density oligo microarrays (Wang *et al.*, 2009). Bacterial transcriptomes have different features from the eukaryotic transcriptomes, in bacterial genomes the neighbouring genes usually overlap; consequently, interfering between the start of one gene transcript from the end of neighbour gene adds complexity to transcriptome analysis, this complexity of starting synthesis of protein which require many other proteins and regulatory elements, RNAP-associated proteins generality affect the processivity of RNAP (McClure *et al.*, 2013). The most complicated challenge to investigate proteome of microbe within a host model is how to recover a sufficient number of bacterial cells from infected tissue (Bumann, 2010, Windle *et al.*, 2010).

1.6.2 Bacterial proteins transcription: genes, promoters and regulatory factors

Different promoters may control expression of genes or operons under specific conditions, and gene models for eukaryotic RNA genes are not similar for bacterial small regulatory RNAs (McClure *et al.*, 2013). Bacterial genes are arranged in operons, which are defined as group of adjoining genes regulated through a single operator (Jacob and Monod, 1961), nevertheless this definition no longer convenient to different gene regulation aspects, it has been thought that this organization leads to an equal level of expression for all genes, however irregular gene expression level within the same operons was proved due to the failure of expression downstream genes in an operon because of the mutation in upstream of the operon and that lead to operon polarity (Güell *et al.*, 2009). Bacterial transcription performed by single RNA polymerase (RNAP) holoenzyme complexes which employ a set of enzymes and σ-factors, the RNAP holoenzyme binds to various σ-factors that recognize different promoters and lead to control specific sets of genes, bacteria have varied numbers of σ-factors, while most of them have at least one of the housekeeping σ-factor family (Paget and Helmann, 2003). The start of protein translation usually requires a short sequence Shine-Dalgarno motif, which is located close to the start codon that recruits the ribosome to the mRNA (Güell *et al.*, 2011).

1.6.3 Transcriptional regulators in *S. aureus*

Proteomics and transcriptomics are powerful combination to investigate genes expression. The relatively small size of *S. aureus* genome and their adaptability for evolution suggests that this bacterium has a high level of genome flexibility, depending on their environment (Holden *et al.*, 2004). Many *S. aureus* strains have been extensively studied, however, the function of great number of their genes is still unidentified. *S. aureus* produces a numerous of virulence factors in a response to environmental changes, this successful behaviour strongly correlated with the regulation of genes expression, RNAs are known as the main regulatory molecules for the physiological and virulence factors of prokaryotes in response to various environmental and stress changes (Beaume *et al.*, 2011). Almost 200 transcripts and antisense RNA molecules were identified and characterized to act as regulatory RNAs by showing that 10% of the intergenic regions in the genome of *S. aureus* contain expressed transcripts (Marie *et al.*, 2010). *S. aureus* encodes 135 transcriptional factors (TFs) and sigma factors were identified and classified into 36 regulatory families, 43% of them have been experimentally characterized (Ibarra *et al.*, 2013).

2.7 Project aims and objectives

The investigations were undertaken with the following aims:

The medical impact of *S. aureus* has led to the development of many concepts about their virulence features. The main aim of this research was to characterise the genetic features of *S. aureus* lipoproteins and examine their expression under different conditions to determine whether they may play a role in the virulence of the bacterium.

- A bioinformatics analysis would be performed to investigate the total number of lipoproteins in *S. aureus* and identify their common molecular features between different strains that had their DNA sequenced. The initial part of this study would be to sequence lipoprotein genes to confirm that those genes were actually present in the examined strains.
- A second aim was to evaluate and identify of lipoprotein genes of *S. aureus* and their expression *in vitro* and *in vivo* conditions, also whether they may play roles in pathogenicity by real-time PCR assay and RNA-Sequencing.

  Finally, the impact of proteomics and transcriptomics tools would be examined which would be used to follow the changes in lipoprotein gene expression profiles in response to *S. aureus* infection in a model system and explore new information on the pathogen physiology and pathogenicity. The host model *C. elegans* would be used to determine the changes in expression of lipoprotein genes and genes encoding different virulence factors that are regulated at different times during the infection process.

- It was hoped that identification of different *S. aureus* lipoproteins and characterising them would clarify their importance for future studies that aim to clarify the molecular basis of the interaction between the invading bacteria and the host, in the field of inflammation, host innate immunity, infectious diseases and development of novel antibiotics and vaccines.

Chapter two

Material and methods

2. Material and methods

2.1 Bacterial strains and growth conditions

The highly successful epidemic hospital-acquired methicillin-resistant *S. aureus* strain EMRSA-16 clone (MRSA252) that was used in these experiments was obtained from the Health Protection Agency culture collections, UK; NCTC13277. *S. aureus* 8325-4 was taken from Professor Howard Foster culture collection. The clinical isolates T1 and RN4282 were taken from our culture collection and were originally supplied by Professor Valerie Edwards-Jones. T1 (FRI189S) was isolated from a case of menstrual TSS (MTSS) and was originally supplied by MS Bergdoll. Strain RN4282 was supplied by T.J. Foster, University of Dublin, Republic of Ireland, and originally obtained from Professor R.P. Novick. All strains were grown in tryptone soya broth media (Oxoid, Basingstoke, UK). Methicillin-resistant *S. aureus* NCTC 12493 was included as methicillin resistant quality control organisms obtained from Culture Collections Public Health England, UK.

2.2 Oxacillin/cefoxitin disc diffusion test

Bacterial sensitivity test was carried out on pure cultures of *S. aureus* strains to determine susceptibility and resistance to methicillin resistance in *S. aureus*, zones of inhibition were measured after incubation for 18-20 h.

2.2.1 Photometric standardization of turbidity of suspensions (Moosdeen *et al.*, 1988)

A few colonies of each tested strain were mixed with 3 ml sterile distilled water to make a bacterial suspension with visible turbidity in a range >0.3-0.6 according to absorbance reading at 500 nm, this suspension was used within 15 min of preparation, 20 µl of bacterial suspension mixed with 5 ml sterile distilled water was used for inoculation of Iso-Sensitest agar plates.

2.2.2 Medium and inoculum preparation

Inoculum was spread evenly over the entire surface of Iso-Sensitest agar (Oxoid, Basingstoke, UK), plates were allowed to dry for 10 min before applying discs of 30 µg cefoxitin, plates were incubated at 35°C in air for 18-20 h. *MecA* positive and methicillin-resistant *S. aureus* NCTC 12493 was included as methicillin resistant control organisms. Cefoxitin test results interpretation were analysed according to the breakpoints of Clinical

& Laboratory Standards Institute, cefoxitin MIC breakpoints for determined inhibition zone was: susceptible = ≥ 22 mm diameter, resistant = ≤ 21 mm diameter.

2.3 Proteomic analysis of *S. aureus* lipoproteins

2.3.1 Lipoprotein extraction

Lipoprotein isolation and fractionation of bacterial cell wall was performed according to the method described by Hashimoto *et al*. (2005). Bacterial growth was monitored spectrophotometrically by measuring at OD_{600} using Camspec M330 UV Visible Photospectrometer (Spectronic Camspec Ltd, UK). Briefly, 1 ml of bacterial cultures OD_{600} of 0.5 was diluted into 100 ml of tryptone soya broth (Oxoid, Basingstoke, UK) in a 250 ml conical flask and incubated at 37°C with constant shaking at 150 rpm. Cultures of late exponential growth phase were harvested after 15 h (Kurokawa. *et al.*, 2012). Bacteria were harvested by centrifugation at 12,000xg for 30 min, final pellets were resuspended in 10 ml of TS buffer (NaCl, 50 mM, Tris HCL10 mM) supplemented with 1 mM protease inhibitor Pefabloc SC (Sigma-Aldrich, UK) to prevent protein degradation. Bacterial suspension was sonicated five times using a tip-probe sonicator for 10 sec and kept in ice to avoid extra heating, the mixture was centrifuged at 10,000xg for 10 min at 4°C to remove insoluble materials then 0.9 ml of bacterial suspension was mixed with 0.1 ml of 20% non-ionic detergent Triton X-114 (Fisher Scientific, UK) and incubated for 2 h at 4°C on a rotator. Phases were allowed to separate at 37°C for 15 min in a water bath followed by centrifugation at 10,000g for 5 min for phase separation (upper aqueous phase and lower detergent phase). The lower detergent phase was collected and washed 3 x with TS buffer and the lipoproteins were recovered by adding 9 times the sample volume of acetone and storing at -20°C overnight. Final pellets were recovered by centrifugation at 13,000xg for 10 min at 4°C. Proteins were quantified by NanoDrop 1000 spectrophotometer (Thermo Scientific) with absorbance at 280nm.

2.3.2 Lipoprotein separation

2.3.2.1 Analytical and preparative 1-D/2-D PAGE

In order to examine the extracted proteins by SDS-PAGE, samples were suspended in Laemmli buffer x 2 (0.5 M Tris-HCl, pH 6.8, glycerol, 12% (w/v) SDS, 0.1% (w/v) bromophenol blue), then loaded in separating gel 12% and stacking gel 4% as in table 1,

by loading 12 µl of each sample and 5 µl of unstained protein molecular weight marker, gels run at 180 V for 2h with SDS-PAGE running buffer (0.192 M glycine, 25 mM Tris-base and 0.1% SDS), the gels were run in mini-protein III gel electrophoresis equipment (Bio-Rad, UK).

Gels were stained for 2h in 45 ml of methanol, 0.1 g coomassie brilliant blue, 100 ml acetic acid and 70% distilled water, then gels were destained in a solution of 100 ml methanol, 35 ml acetic acid in 1L distilled water, with shaking for 3h.

Table 1. SDS-PAGE loading gel concentrations

Solution	Separating gel 12%	Stacking gel 4%
Sterile deionised water	9.6 ml	4.96 ml
30% acrylamide mix	8 ml	0.74 ml
0.4% SDS in 1.5 M Tris-base pH 8.8	6 ml	-
0.4% SDS in 0.5 M Tris-base pH 6.8	-	1.86 ml
10% ammonium persulphate	90 µl	75 µl
Tetramethylethylenediamine	9 µl	9 µl

2.3.2.2 Protein separation in 2-D gels

Complex proteins samples from bacterial cells can be separated to produce an individual protein spots to give clear 2-D maps of the proteome. First dimension was carried out using two different sizes pH 3-10 18-cm I and pH 3-5.6 7 cm PG non-linear Immobiline DryStrip gel strips (Amersham Pharmacia Biotech, Sweden). Approx. 100 µg proteins prior to analysis were rehydrated in 340 µl of rehydration buffer [7 M urea (Sigma-Aldrich, UK), 2 M thiourea (Sigma-Aldrich, UK), 2% CHAPS (Sigma-Aldrich, UK), 0.5% Bio-Lyte pH 3-10 carrier ampholytes, 0.2% DTT and 0.1% bromophenol blue solution]. Amersham Ettan™ IPGphor II isoelectric focusing system (GE Healthcare, UK) was used. IPG gel strips were isoelectrically focused for a total of 85000 Vh, running conditions were as follows: 500 V for 500 Vh, followed by two gradients of 1000 V for 800 Vh and 8000 V for 13 500 Vh and finally 8000 V for 20 000 Vh. IPG gel strips were equilibrated in 75 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 2% SDS and 0.1% bromophenol blue solution, including 1% DTT in the reduction step (15 min) and 2.5% iodoacetamide in the alkylation step (15 min). The second dimension was performed on 12.5% SDS-

polyacrylamide slab gel (25 x 19.5 x 0.15 cm) was prepared in Multi-Casting Chamber (Bio-Rad), strips were embedded in 0.5% agarose in cathode buffer (0.192M glycine, 25 mM Tris-base and 0.1% SDS). Separation was carried out at room temperature, gels were run by electrophoresis in a Protean II xi 2-D Cell (Bio-Rad) at 110 V for the first 20 min and then at 180 V until the bromophenol blue reached the bottom of the gel. Protein spots in all gels were stained for 1 h with coomassie brilliant blue solution, gel was destained in a solution of 100 ml methanol plus 35 ml acetic acid in 1L distilled water for 3 h, stained gels were scanned using a Quantity One scanner (Bio-Rad). Individual spots were excised from gels and subjected to trypsin digestion.

2.4 In-gel Trypsin digestion

Stained proteins spots and bands containing target proteins were subjected to series of reduction and alkylation reaction in Eppendorf tubes, the alkylation reaction results in carbamidomethyl modified cysteine residues. The first step was reduction of the gel spots with 50 mM ammonium bicarbonate $NH_4 HCO_3$ (Fisher Scientific, UK), 150 µl of $NH_4 HCO_3$ was added to each gel spot and incubated for 15 min then the ammonium bicarbonate was discarded and replaced with 150 µl of 50% acetonitrile for each tube to remove coomassie dye and incubated for 15 min, this process was repeated for three times to remove all stain, 5 mM of dithiothreitol (Fisher Scientific, UK) was added to each tube and incubated for 1 h in a heating block at 60°C, in this step proteins were subjected to denaturation and alkylation of disulfide bonds at the cysteine residues. After removing the supernatant, 15 mM iodoacetamide (Fisher Scientific, UK) was added to each tube and incubated for 1 h in the dark, this step prevent proteins to form disulfide bonds by binding iodoacetamide which act as irreversible inhibitor of all cysteine peptidases. Supernatant was removed and replaced with 50 mM $NH_4 HCO_3$ and incubated for 15 min at room temperature, liquid was discarded and 150 µl of 50% acetonitrile added and left for another 15 min, after this step as much liquid as possible was removed. Trypsin Sequencing Grade enzyme (Promega, Madison, WI, USA) which specifically hydrolyzes peptide bonds at the carboxylic sides of lysine and arginine residues, solution at a ratio of 1:100 (20 µg was mixed with 20 µl of a stability-optimized resuspension buffer solution and 400 µl of 50 mM $NH_4 HCO_3$, this mixture were added to gel spots tubes and incubated at 37°C for at least 8 h then the liquid was collected and kept in new tubes labelled with final peptide digested tube, while 100 µl of acetonitrile TFA solution was added to gel spots and

incubated at room temperature for 1 h and the final supernatant were added to final peptide digested tubes, extra extraction of liquid from the digest gel was extracted by addition of 40 µl of water, extracted tryptic peptides were analysis by mass spectrometry.

2.5 Protein identification by LC/MS/MS

Tryptic peptides were desalted and purified by using an in-house manufactured C18 purification tip, briefly, C18 Zip-Tip (10 µl pipette tip with 0.5 µl bed of C18 silica based medium fixed at its tip) was wetted with 15 µl of 50% acetonitrile and 0.1% TFA, C18 tip was equilibrated with 15 µl of 0.1% TFA, material was slowly aspirated and dispensed, desalting was performed by aspirating and dispensing 10 µl of 0.1% TFA, then peptides were eluted with 20 µl of 50% acetonitrile and 0.1% TFA, then peptides were analysed on Thermo QExactive mass spectrometer coupled to a Dionex RSLC nano system (Dionex), samples were resolved on a Nano EASY (Thermo Scientific, UK) C18 reverse-phase column fitted with a 50 cm long, 75 µm internal diameter, reverse-phase column and resolved using a 60 min gradient. Liquid chromatography and mass spectrometry was performed at Central Proteomics Facility, Sir William Dunn Pathology School, Oxford University. A nano-spray injection was employed to introduce samples into a Brucker Esquire 3000 plus ion trap mass spectrometer. Generated raw data file (.dat) was analysed by data processing MaxQuant software package to create peak lists mascot generic file (.mgf) based on the acquired MS/MS spectra, proteins were identified by Mascot server (version 2.4; Matrix Science, London, UK), outcome data were searched against the UniProt KB/Swiss-Prot of *S. aureus* MRSA252 proteins sequence database.

2.5.1 Protein identification by peptide mass fingerprinting/MALDI TOF-MS

Tryptic peptides obtained from gel spots digestion were mixed with a solution of 50% acetonitrile, 0.1% trifluoroacetic acid (Sigma-Aldrich, UK) and 10 mg α-cyano-4-hydroxycinnamic acid /ml (Sigma-Aldrich, UK) and then placed onto 96 well 1/4 microtitre stainless steel plate and incubated for 10 min to crystallize. Samples were analysed by MALDI-TOF MS operated in positive linear mode, internal calibration containing mixed of four peptides are summarized in table 2, were used to calibrate results. A brief summary in figure 7 shows the proteomic outline for all methods used to investigate lipoprotein in *S. aureus* including gel based and gel-free peptide digestion procedures.

Table 2. Mixture of 20 µM stock solutions made up of 4 peptides of known sizes used as internal calibration

Peptides	Size of peptide	70 µl solutions contain
Bradykinin fragment 2-9	904.02 g/mol	10 µl
Bombesin (acetate salt hydrate)	1,619.8 g/mol	10 µl
Angiotensin II (acetate)	1,032.1 g/mol	20 µl
ACTH (adrenocorticotropic hormone fragment 18-39 human)	2,465.6 g/mol	30 µl

2.5.2 Quantitative proteomics techniques

2.5.2.1 In-solution digestion, Liquid Chromatography and Mass Spectrometry (LC/MS/MS)

One hundred micrograms of protein mixtures from each sample were trypsin digested as described in section 2.4 section and the resulting digest was analysed by the LC/MS/MS, analysis performed using a Dionex RSLC nano ultrahigh pressure HPLC system fitted with a 50cm long and 75 micron internal diameter reverse-phase column and resolved using a 60 min gradient and coupled with high-resolution Thermo Q-Exactive Orbitrap Mass Spectrometer Liquid chromatography, the mass spectrometry and proteins identified by Mascot were performed at Central Proteomics Facility, Sir William Dunn Pathology School, Oxford University.

The raw data files were analysed by data processing quantitation MaxQuant software package, version 1.2.2.5 (Cox and Mann, 2008), to create peak lists mascot generic file, then proteins identified by Mascot (version 2.4; Matrix Science, London, UK), tryptic digested peptides generated from mass spectrometric were searched against proteins database followed by matching with relatively corresponding proteins list of UniProt KB/Swiss-Prot of *S. aureus* MRSA252 proteins sequence. Spectra were normalised by the method of spectral abundance factor (NSAF) to calculate samples variations of expression changes (Zybailov *et al.*, 2006).

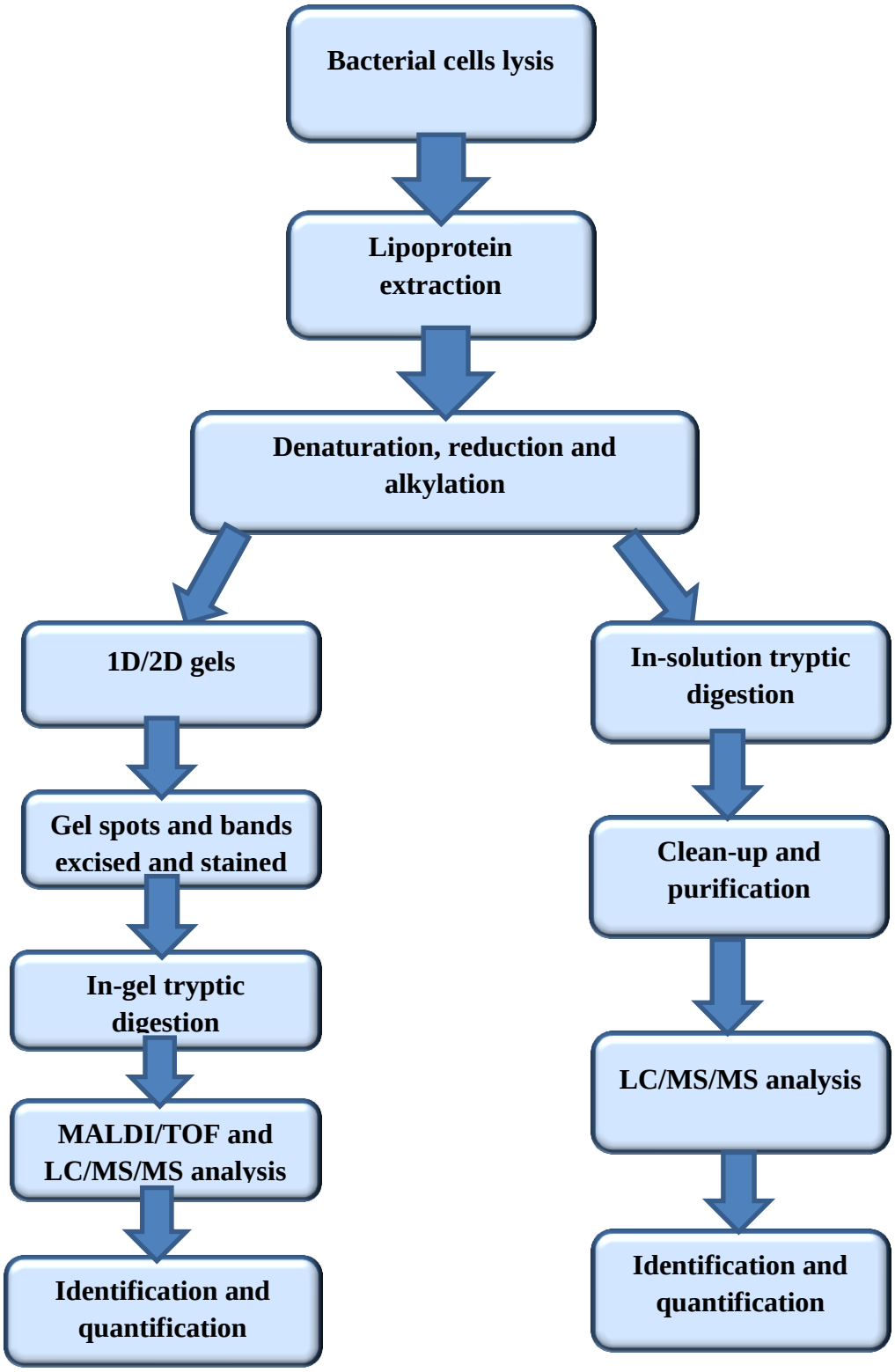

Figure 7. Workflows of in-gel and in-solution protein digestion procedures

In-gel (left) and in-solution (right) protein digestion procedures and subsequent LC-MS/MS analysis of lipoprotein samples.

2.6 Bioinformatics of *S. aureus* genomic DNA

2.6.1 Bacterial genomic DNA extraction

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) as in the manufacturer's protocol. Briefly, harvested bacterial pellets were washed twice with 0.15 M NaCl, 10 mM EDTA (pH 8.0) and resuspended in lysozyme solution (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 10 mg/ml lysozyme and 1mg/ml lysostaphin) and incubated for 30 min at 37°C, lysed cells were added to 25 µl proteinase K (Qiagen) and 200 µl AL buffer, samples were mixed and incubated at 56°C for 30 min, followed by adding 200 µl 100% ethanol and vortexed to homogenise the solution which was transferred into DNeasy Mini spin columns and centrifuged for 1 min at 6,000 x g. Columns were placed in new 2 ml collection tubes, 500 µl AW1 buffer were added and centrifuged for 1 min at 6000 x g, finally 500 µl AW2 buffer were add and centrifuged for 3 min at 17,000 x g prior to eluting DNA with 100 µl TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA), DNA was verified by 1.5% agarose gel electrophoresis, isolated DNA stored at -20°C till further use.

2.6.2 Polymerase Chain Reaction (PCR)

All primer pairs were designed by Primer-BLAST (Ye *et al.*, 2012), on the basis of published *S. aureus* genome MRSA252 (GenBank accession no NC_002952), by alignment of published DNA sequences of target lipoprotein genes at http://www.ncbi.nlm.nih.gov/ tools/primer blast/ index. *S. aureus* MRSA252 strain list of lipoprotein genes examined in this experiment with the accession numbers, oligonucleotide primers, nucleotide location and size of PCR product for each gene fragments are given in table 4.

PCR amplification was performed in a thermal cycler (Stratagene Robocycler Gradient 96 Thermal Cycler), VELOCITY DNA Polymerase (Bioline, UK) was used, reaction mixture consisted of 25 µl, 5 µl 5x Hi-Fi reaction buffer, 1 µl 10 mM dNTP Mix, 1 µl 10 µM each primers, 1.5 µl DMSO, 1 µl polymerase enzyme, 20 ng DNA template and 15 µl of nuclease free water. PCR reaction mixtures were heated for initial denaturation at 94°C for 2 min; followed by 35 cycles of amplification; denaturing, annealing and extension at 94°C (30 s), 58°C (1 min) and 72°C (1 min); with one final extension cycle at 72°C for 10 min. PCR products were separated by 1.5% agarose gel electrophoresis in 1x TBE buffer

(pH 8; 0.09 M Tris, 0.09 M boric acid, 2 mM EDTA) and 0.03% (wt/vol) GelRed (Biotium, UK), incorporated for DNA staining. Five microliters of PCR products were mixed with 5 µl of loading buffer and applied to each gel well; gels were run in 1x TBE buffer at 100 V for 1 h then visualized and photographed on a UV transilluminator. A 50 bp Hyperladder II (Bioline, UK) was used as molecular size markers in all gels.

2.6.3 *S. aureus* genes sequencing, annotation and comparative analysis

PCR products were purified by using the QIAquick PCR purification kit (Qiagen) and subjected to Sanger sequencing technology (Sanger *et al.*, 1977) at Eurofins MWG Operon. Sequencing was performed using either of the primers used for PCR, forward and reverse reads were generated for all amplicons and analysed using FinchTV viewer (Version 1.4.0, Geospiza, Inc., Seattle, WA, USA; http://www.geospiza.com). Each chromatogram file of nucleotide sequences for complete sequenced lipoprotein genes of RN4282 and T1 *S. aureus* strains were individually checked for final use by using FinchTV viewer, while translated predicted lipoprotein DNA sequences for other 18 strains were obtained from publicly available databases the National Centre for Biotechnology Information (NCBI), http:// www. ncbi.nlm.nih.gov. Table 3 shows the *S. aureus* strains used in BLAST comparative analysis. Sequence homology of each target gene in a genome was first identified using BLAST function of GenBank database http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM =blastn &PAGE_TYPE= BlastSearch & LINK_LOC=blasthome, a list of predicted lipoprotein and their amino acids sequences of *S. aureus* MRSA252 is available at http://www.mrc-lmb.cam.ac.uk/ genomes /dolop /predicted/a.shtml, and included in section 3.2.4 (table 12). All sequence reads were aligned against *S. aureus* MRSA252 strain, 50 lipoprotein selected genes were sequences and aligned to identify SNPs and large regions of difference, nucleotides sequences were trimmed with Proseq 3.5 (Filatov, 2002).

Table 3. *S. aureus* strains used in BLAST comparative analysis

***S. aureus* strain**	**Pathogenicity**	**Host**	**Reference**
S. aureus MRSA252	HA-MRSA	Human	(Holden *et al.*, 2004)
S. aureus RN4282	MSSA	Human	(Edwards-Jones and Foster, 2002)
S. aureus T1	MSSA	Human	(Edwards-Jones and Foster, 2002)
S. aureus COL	HA-MRSA	Human and animals	(Gill *et al.*, 2005)
S. aureus MSSA476	MSSA	Human and animals	(Holden *et al.*, 2010)
S. aureus MW2	CA-MRSA	Human	(Baba *et al.*, 2002)
S. aureus N315	HA-MRSA	Human	(Kuroda *et al.*, 2001)
S. aureus Mu50	HA-MRSA	Human	(Kuroda *et al.*, 2001)
S. aureus T0131	HA-MRSA	Human	(Li *et al.*, 2011)
S. aureus TW20	HA-MRSA	Human	(Holden *et al.*, 2010)
S. aureus BMB9393	HA-MRSA	Human	(Costa *et al.*, 2013)
S. aureus LGA251	MRSA	Human and animals	(García-Álvarez *et al.*, 2011)
S. aureus RF122	MSSA	animals	(Herron-Olson *et al.*, 2007)
S. aureus ST398	MRSA	Human and animals	(Witte *et al.*, 2007b)
S. aureus M013	CA-MRSA	Human	(Huang *et al.*, 2012)
S. aureus M1	MRSA	Human	(Larner-Svensson *et al.*, 2013)
S. aureus Mu3	HA-MRSA	Human	(Neoh *et al.*, 2008)
S. aureus str. Newman	MSSA	Human	(Baba *et al.*, 2008)
S. aureus NCTC8325	MSSA	Human	(Iandolo *et al.*, 2002)
S. aureus VC40	VRSA	Human	(Sass *et al.*, 2012)

Table 4. Oligonucleotide primers, base sequences, genes locations, and predicted sizes of PCR products for *S. aureus* lipoproteins

ORF indicates the gene locus in MRSA252 strain (http://www.genedb.org/Query/geneType?taxons=*Saureus*_MRSA252&type=mRNA)

Gene ORF	Gene accession No	Primers	Oligonucleotide sequence (5'→3')	Nucleotide location within genome (bp)	Size of PCR product (bp)
SAR0118	YP_039582	Forward Reverse	TGTTTTTCAATATTTAACTTTTCATA GAATAAAGTAATTAAAATGCTTGTTGTTA	128,926..129,918	993 bp
SAR0174	YP_039639	Forward Reverse	ATGAAAAGGTTAAGCATAATCGTCA TCATGAACGTGATGCCTCCT	191,762..192,736	975 bp
SAR0201	YP_039666	Forward Reverse	GTGAAGAAAATCATTAGTATCGCA TTATTGATTAATGGCTTTTTCTACT	231,906..233,681	1776 bp
SAR0206	YP_039671	Forward Reverse	TAAAATATATCACGTTAGCCGTGGTAATG TTTTGTGACGGATGAAGAATCTTAA	239,168..240,439	1272 bp
SAR0216	YP_039681	Forward Reverse	TCATTGAATCATCTCCAAAAA GAAATCAAAAATTTATATCTTGCTA	251,020..251,988	969 bp
SAR0230	YP_039695	Forward Reverse	AGTTCGATAATCGATTAAATAGATACCT TGAAGTTTAAAAGACTAGCAACTATATT	271,601..273,076	1476 bp
SAR0340	YP_039797	Forward Reverse	ATGAAAAAGTTAACAACGCT TCATTCAGTAATCACAGCCA	385,032..385,886	855 bp
SAR0201	YP_039666	Forward Reverse	GTGAAGAAAATCATTAGTATCGCA TTATTGATTAATGGCTTTTTCTACT	231,906..233,681	1776 bp
SAR0340	YP_039797	Forward Reverse	ATGAAAAAGTTAACAACGCT TCATTCAGTAATCACAGCCA	385,032..385,886	855 bp
SAR0390	YP_039845	Forward Reverse	ATGAAATTAAAATCATTAGCAGTGT TTATTGATCTTGCTCACTCT	426,962..427,534	573 bp
SAR0396	YP_039852	Forward Reverse	TTATTTATCGATAACATCACTCTTGA AGAGATTACTACTAAGTACATTTTTAGCA	430,210..430,836	627 bp
SAR0438	YP_039889	Forward Reverse	ATGATGGGAAATATAAAAAGTT TTAGCTATCTTCATCAGACG	467,513..468,295	786 bp
SAR0439	YP_039890	Forward Reverse	AGGTTTGCATTGTACATAAGCG TGGTTTATAACTTAAAAATTCAATGA	468,343..469,116	774 bp

Table 4-continued

Gene ORF	Gene accession No	Primers	Oligonucleotide sequence (5'→3')	Nucleotide location within genome (bp)	Size of PCR product (bp)
SAR0442	YP_039891	Forward Reverse	ATGGGATATTTAAAAAGGATTG CTATTTTTTCGCTGGCTTAT	469,987..470,757	771 bp
SAR0443	YP_039892	Forward Reverse	ATGAGATATTTAAATAGAGTTGTACTGTAC TCATTCATCATCTTCATTCACC	470,789..471,589	801 bp
SAR0444	YP_039893	Forward Reverse	TGAAGTCTATAAAAAGGATTGGAT TTAGTAATTATATTTATCCTCGCTT	471,608..472,402	794 bp
SAR0463	YP_039912	Forward Reverse	ATGAAAAGATTGATTGGGTTAGTTATC TTATTTTGCTAATGACGTTTCTATTGCC	491,739..492,581	842 bp
SAR0618	YP_040063	Forward Reverse	GTGAAGAAATCGTTAATTGCT TTATTTTCTATAAATTGCATCTCT	666,806..667,693	888 bp
SAR0641	YP_040083	Forward Reverse	TTATTTCATGCTTCCGTGTA AAAAATTAGTACCTTTATTATTAGC	687,790..688,719	930 bp
SAR0706	YP_040140	Forward Reverse	TTATTTTTTATCACGTTTATGGAGCC GTGCGATTTATGAATGAAATTCTTGTT	746,977..747,084	105 bp
SAR0730	YP_040157	Forward Reverse	ATGAAGAAATTAATCATCAGC TTATGATTGGTGTTTGTCATTA	766,363..766,752	390 bp
SAR0790	YP_040217	Forward Reverse	ATGAAGAAAACAGTCTTATATTTAGTA TTATTTTACAACTTTTTCAAGTTC	828,218..829,246	1029 bp
SAR0794	YP_040221	Forward Reverse	ATGAAAAAAATTGTTATTATCGC TTATTTTTCTTCTAATTTTTCAAGC	831,583..832,461	879 bp
SAR0872	YP_040295	Forward Reverse	AATTATTTGGTCTTATTTTAGTATTA CACCGTTGTATTTTTCATTAA	910,328..911,149	822 bp
SAR0953	YP_040375	Forward Reverse	ATGACAAGAAAGTTAAAAACGCTGAT TTATTTTTTCTTCTTACCTGTTTCTTTA	995,948..997,603	1656 bp
SAR1011	YP_040427	Forward Reverse	GTGAATAGGAATATCGTTAAATTAGTTGT TAAATCTTCTGCCATACTTTCACTTG	1,054,214..1,055,173	960 bp

Table 4-continued

Gene ORF	Gene accession No	Primers	Oligonucleotide sequence (5'→3')	Nucleotide location within genome (bp)	Size of PCR product (bp)
SAR1034	YP_040449	Forward Reverse	TTAATGTCCACCTCCATGA GTGTCAAAATTTAAGTCTTTGCTT	1,079,932..1,081,032	1101 bp
SAR1066	YP_040479	Forward Reverse	ATGAAATTTGGAAAAACAATCGCA TTATTTAAATTGATCAACGTCTTGC	1,112,113..1,112,739	627 bp
SAR1106	YP_040519	Forward Reverse	TTGAGAATCATAAAGTATTTAACCATTT AATCATATAATTGAGTCATTGCCTTATC	1,152,065..1,152,943	879 bp
SAR1189	YP_040600	Forward Reverse	ATGAAAAAGACACTGGGATGT TATTTTAAAGAATCATCTGACGCTG	1,238,170..1,239,123	954 bp
SAR1288	YP_040698	Forward Reverse	TCATATTATTTTCTTCATAAACTGG TGAGGCGATGGTTTGTATT	1,352,458..1,352,814	357 bp
SAR1494	YP_040896	Forward Reverse	GTCAAACACACCATATTCTGA TGTTTAAAAGAACTAAACTAATCTT	1,589,460..1,590,368	905 bp
SAR1495	YP_040897	Forward Reverse	TCATTCGACCTCAATCCTTATAGAC GGCAAAATTAATCTTAATAGCAACG	1,590,426..1,591,331	906 bp
SAR1558	YP_040960	Forward Reverse	AAAAAGTAATCGGACTGCTAC TTAATATTGGCTCCTGGTACTG	1,634,418..1,634,852	427 bp
SAR1608	YP_041004	Forward Reverse	ATGAAAAAATTGGTTTCAATTGTTGGC TCATGATTTTGCATTTAAGTTTAATTTTG	1,681,001..1,681,582	582 bp
SAR1831	YP_041217	Forward Reverse	AGTTAATATTTTTAATTGTAATTGCTTTA AATTCCTTCATTACACTCTTGGC	1,913,827..1,914,672	846 bp
SAR1879	YP_041264	Forward Reverse	ATGTTAAAAGGATGCGGCGG ATTTATCAGGTTCACATGCACGTTT	1,966,068..1,966,682	555 bp
SAR1881	YP_041266	Forward Reverse	AATTCAAAGCTATCGTTGCAATCAC TAGTTTCTGCTGAATTATTTTCACG	1,967,834..1,968,460	627 bp

Table 4-continued

Gene ORF	Gene accession No	Primers	Oligonucleotide sequence (5'→3')	Nucleotide location within genome (bp)	Size of PCR product (bp)
SAR1995	YP_041368	Forward Reverse	TTAATTACTGTAAATATGAACTTGCGG ATGAAGCGTACATTAGTATTATTGATTA	2,084,777..2,085,976	1200 bp
SAR2104	YP_041470	Forward Reverse	AGATTGTTAGGTTTATTATTAGTGAGCAC TCAATTATCATTATTTATAATTTCAGAAA	2,166,873..2,167,337	460bp
SAR2179	YP_041538	Forward Reverse	TTATTTCTTTTTCTTTTTAGACACTACT ATGAAGAAAAAAACGTTACTACCATTAT	2,249,989..2,250,861	873 bp
SAR2268	YP_041621	Forward Reverse	TTACTTTTGTTCTTTTTTTGATAAT ATGAGAGGTCTAAAAACTTTTAGTATAT	2,349,338..2,350,321	984 bp
SAR2363	YP_041718	Forward Reverse	TTATGCTGTAAAGTGGTATTCTT ATGAAAATGAAACGTTTTATAGCT	2,433,026..2,433,808	783 bp
SAR2368	YP_041723	Forward Reverse	TTGCAGCTTTAATTAATTTTTCTT ATGAAAAAACTATTATTACCATTAATAA	2,437,923..2,438,831	904 bp
SAR2457	YP_041810	Forward Reverse	TTATTGTTGGTAGTTTGGATCAG AAAATTAGTTACAGGGTTATTAGC	2,528,101..2,528,730	630 bp
SAR2496	YP_041846	Forward Reverse	TTAATGCGCTAACATTTCTTCT ATGAAAAAGAAATTAGGTATGTTACTT	2,569,326..2,570,873	1548 bp
SAR2499	YP_041849	Forward Reverse	TTGATTTTATCTTTTAATAATTTTTCATA AATTACTAACATTATTTATAGTGAGCATG	2,571,954..2,572,553	600 bp
SAR2500	YP_041850	Forward Reverse	CTAGTTCGTCATATTTTCTTCAT TTTGTGGCGACGGTATTATT	2,572,572..2,572,934	363 bp
SAR2504	YP_041854	Forward Reverse	AGATTTAGAAACATCTTGACCAAAC AAAGACTTTTATTTGTGGTGATAGC	2,575,960..2,576,739	780 bp
SAR2536	YP_041887	Forward Reverse	TTACTTATGACCACCTTTCTGTTTAT AATATATACTTGTCGTGCTTGTCTTA	2,615,364..2,616,305	942 bp
SAR2546	YP_041897	Forward Reverse	GAAAAAATTATGTTCATTAATTGTAGTAG CTATTTGTCGTCCATTTTATCTTTT	2,627,660..2,628,115	456 bp
SAR2554	YP_041905	Forward Reverse	CATTTCATTGAATGGTAATTCATACTGT ATGAGAAAACTAACTAAAATGAGTGCAAT	2,635,602..2,637,200	1599 bp

2.6.4 Computation of sequence parameters

Ratios of synonymous and nonsynonymous nucleotide substitutions were calculated using the method of Nei & Gojobori, 1986, which is implemented in MEGA version 6 (Tamura *et al*., 2013). Nucleotide diversity (π) the mean percentage of polymorphic sites over all pairwise comparisons was calculated by MEGA version 6. The ORF (Open Reading Frame) were identified by using ORF finder http://www.ncbi. nlm.nih.gov/gorf, to display evidence of frame shifts or mutations leading to a premature stop codon, nucleotide diversities and species divergence calculations were performed using MEGA 6 and DnaSP v5 (Librado and Rozas, 2009). Calculations of nucleotide (*n*) diversities for each gene were determined by MEGA 6 and DnaSP v5 software. Nucleotide diversity in synonymous, nonsynonymous and silent sites (synonymous sites are those sites in a codon where nucleotide changes result in synonymous substitutions) were computed using method of Nei and Gojobori 1986.

2.6.5 Phylogenetic analysis and the concatenation approach

There are various ways to construct phylogenetic trees from multiple genes for the same set of species. The most recent phylogenetic tree reconstruction is the genes concatenated head-to-tail to form a super-gene alignment. The statistical advantage of this method gives more phylogenetic accuracy by increasing the sample size.

Reconstruction of evolutionary relationships was carried out by using two different software packages, Seaview 4 (Gouy *et al.*, 2010), and MEGA version 6 (Tamura, *et al* 2013) to confirm the phylogenetic results. Bayesian phylogeny was reconstructed from the concatenated sequences, 44 genes varying in sequence and in length from 105 to 1776 bp were used for the phylogenetic analysis (six genes were not present in all strains), representing 37670 bp were used to construct trees using the neighbour-joining tree method (Saitou and Nei, 1987), using default parameters except where specified. Nucleotide sequences were aligned using Clustal W2 (Larkin *et al.*, 2007), with a manual adjustment. In order to root all created trees, the mid-point rooting (MPR) method were employed, the root of an unrooted tree is placed at the mid-point of the longest distance between the two most divergent operational taxonomic units (Hess and De Moraes Russo, 2007).

Nucleotide homologous sequences were aligned with clustalW2 software (Larkin *et al.*, 2007), however, phylogenetic analyses were applied to the concatenated nucleotide sequence of 44 lipoprotein genes (the most commen lipoprotein genes) for each *S. aureus* strain, the phylogenetic reconstruction method used neighbor-joining (NJ) method implemented in MEGA 6 software, the main advantages of this method is fast, suited for large datasets and for bootstrap analysis (Felsenstein, 1997). NJ tree was based on the gamma-corrected nucleotide distance, with parameter (a = 0.01) estimated by MrModeltest 2.3 program (Posada and Crandall, 1998). In NJ trees, the reliability of internal branches was assessed by bootstrapping method for estimating the standard error (Efron and Tibshirani, 1986), which is installed in MEGA 6, 1,000 bootstrapping consensus pseudosamples were used. Tree drawing was managed with FigTree application (FigTree, 2014).

2.6.6 Nucleotide substitution and SNP analysis

The generated nucleotide sequences for all genes were analysed to calculate the number of synonymous nucleotide substitutions (*dS*) and nonsynonymous nucleotide substitutions (*dN*) by a maximum-likelihood method using the DnaSP program, the outcome values were consistent with the phylogenetic analysis results. In order to investigate the single nucleotide polymorphisms SNPs, lipoprotein genes sequence of all strains were aligned against MRSA252 *S. aureus* reference genome.

2.6.7 Lipobox features

The publicly available complete genome sequences for *S. aureus* strains provided a useful starting point to research for lipoprotein with the improved multiple sequence alignment programs. The translated protein sequence of 50 predicted lipoprotein genes for two sequenced *S. aureus* and 18 other previously sequences strains available at the National Centre for Biotechnology Information NCBI http:// www. ncbi.nlm.nih.gov, were chosen to examine lipobox features, the first 45 amino acids sequences from each lipoprotein genes were aligned using MEGA 6 software to identify the consensus sequence.

2.7 Isolation of RNA and assessment quality/concentration

2.7.1 RNA extraction

Total RNA was isolated from *S. aureus* using PureLink RNA Mini Kit (Ambion) as in the manufacturer's protocol, with slight modifications as outlined below. Briefly, 3 ml of bacterial culture cells ($\leq 1 \times 10^9$ cells) at two times of exponential phase (4 and 8 h) were harvested and suspended with 6 ml RNAprotect Bacteria Reagent (Qiagen, Hilden, Germany), vortexed for 30 sec and incubated at ambient temperature for 10 min. The mixture was then centrifuged for 10 min at 5,000 x g. Pelleted cells were resuspended in lysozyme solution 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 10 mg/ml lysozyme, 1 mg/ml lysostaphin and Proteinase K, samples were incubated for 30 min at 37°C with shaking, 350 µl lysis buffer prepared with β-mercaptoethanol (10 µl β-mercaptoethanol per 1 ml) was added and vortexed for 1 min. Samples were then homogenized in high-speed homogenizer (Invitrogen) and the supernatant was diluted with 100% ethanol. Binding, washing and elution steps were carried out in a Mini Spin column as in the manufacturer's instruction. Eluted RNA was treated with DNase I, Amplification Grade (Invitrogen, UK), 1 U/ 1 µg of total RNA at 37°C for 60 min to digest single and double stranded DNA, DNase I was inactivated by addition of 1 µl of 25 mM EDTA and the reaction mixture was incubated for 10 min at 65°C, removal of genomic DNA was confirmed by PCR. Integrity of total RNA was carried out at 70 V for 1 h using a 1% (w/v) agarose gel and stained with Gel Red (Biotium, UK), RNA concentrations were measured by NanoDrop 1000 spectrophotometer (Thermo Scientific), three independent measurements of the same sample were performed, all RNA isolations resulted in acceptably low levels of protein contamination (A260/A280 ≥1.95), ratio of sample absorbance at 260 and 280 nm of ~2.0, samples stored at - 80°C for further use.

2.7.2 cDNA synthesis

cDNA was synthesized using Superscript III reverse transcriptase (Invitrogen). One microgram of total purified RNA was combined with 100 ng of random hexamers primers and 1ul 10 mM dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP) at neutral pH, then the mixture was heated at 65°C for 5 min and incubated on ice for 2 min. cDNA synthesis reaction volume consisted of 4 µl 5X first-strand buffer, 1 µl 0.1M DTT, 1 µl RNaseOUT 40U / µl (Ambion) and 1 µl of SuperScript III RT 200U/ µl, 1 µg RNA and 1ul 10 mM of random hexamers primers, tubes was centrifuged briefly and incubated at 25°C

for 5 min, 55°C for 60 min, and reaction inactivated by heating at 70°C for 15 min, prior to degrading RNA with RNase H (Biolabs) for 20 min at 37°C, cDNA purified using a QIAquick PCR purification kit (Qiagen).

2.7.3 Quantitative RT-PCR using dual labelled probes

Lipoprotein gene expressions were determined by TaqMan quantitative RT-PCR to find out the relative expression levels of each gene under investigation within the same growth conditions. High purified salt-free oligonucleotide primers and dual labeled TaqMan probes were designed by using software provided by Eurofins Genetics Services, UK, http://ecom2.mwgdna.com/services/webgist/dual probe_design.tcl?ot=OLIGO DLPD& itemid=new, also synthesized by Eurofins Genetics services are listed in table 5. TaqMan probes were labeled with FAM on the 5' end and TAMARA on the 3' end. Predicted qPCR product sizes were in 100-150 bp range, while all experiments were performed in triplicate. Three independent experiments were conducted for 4 h and 8 h of grown culture, optimized PCR reactions were conducted in a total volume of 25 µl containing 2 µl of cDNA (1 µg of RNA), 12 µl 2 x QuantiTect Probe PCR Master Mix (QuantiTect Probe PCR Kit Qiagen) containing HotStarTaq DNA Polymerase, 1x PCR buffer (Tris Cl, KCl, $(NH_4)_2SO_4$, 8 mM $MgCl_2$, pH 8.7), 0.4 mM each deoxynucleoside triphosphate (dATP, dCTP, dGTP, and dTTP) and ROX Fluorescent dye, 10 pmol of each forward/reverse primers and 1 pmol of dual labeled probe and RNase-free water up to the final volume, reaction mixture was loaded onto semi-skirted PCR plates and snap-sealed by a plastic cap, PCR light cycler reactions were amplified and quantified using DNA Engine Opticon 2 System. Cycling conditions used for all genes were: 1 cycle of initial denaturation at 95°C for 15 min followed by 39 cycles of amplification at 94°C for 15 s, combined annealing, extension at 60°C for 1 min and reading fluorescence at regular intervals after the annealing and extension phases. Alongside a negative control (RNA) a non-template control (NTC) was performed in order to ensure amplification the negative samples was implemented, relative amounts of cDNA were normalized to *S. aureus* 16S rRNA gene as an endogenous control (GenBank accession no. Y15856). Gene expression levels ratio was approximated by two relative quantification normalized to reference gene equations to confirm all results, (i) The Pfaffl method: where ΔCt = Ct (calibrator) – Ct (test) (Pfaffl, 2001).

Ratio = (E target) ΔCT, target (calibrator – test)

(E ref) ΔCT, ref (calibrator – test)

(ii) Livak method: $2^{-\Delta\Delta Ct}$ where ΔΔCt = ΔCt (test) – ΔCt (calibrator) (Schmittgen and Livak, 2008). MRSA252 4 h sample was chosen as calibrator for all samples, thus expression of the tested genes in other samples is expressed as an increase or decrease relative to the calibrator. Error bars represented the standard deviation of the mean expression values.

2.7.4 Quantitative RT-PCR amplification efficiency (E)

One of the main concerns regarding any real-time PCR assay product is the amplification efficiency (E) of RT-PCR experiment, amplification efficiency values of examined genes were measure by a evaluating the doubling of the reaction product at every RT-PCR cycle, Ct slope method (Pfaffl, 2004), by generating a serial dilutions (10 fold dilution) of target template and measuring the Ct value for each dilution. A standard curve for each gene was obtained by plotting threshold cycle Ct on the Y-axis and the log of cDNA concentrations (ng/µl) on X-axis. Amplification efficiency for the tested genes was calculated from the slopes of standard curve using the formula (E = 10–1/slope) when the expected slope value for a 10 fold serial dilution of template is -3.32. The ΔΔCt method was used to calculate RT-PCRs assay results with efficiencies close to 100%, amount of internal reference gene relative to a calibrator (fold change between two Ct values). The amplification efficiency (E) for gene *SAR0216* as example is shown in figure 8.

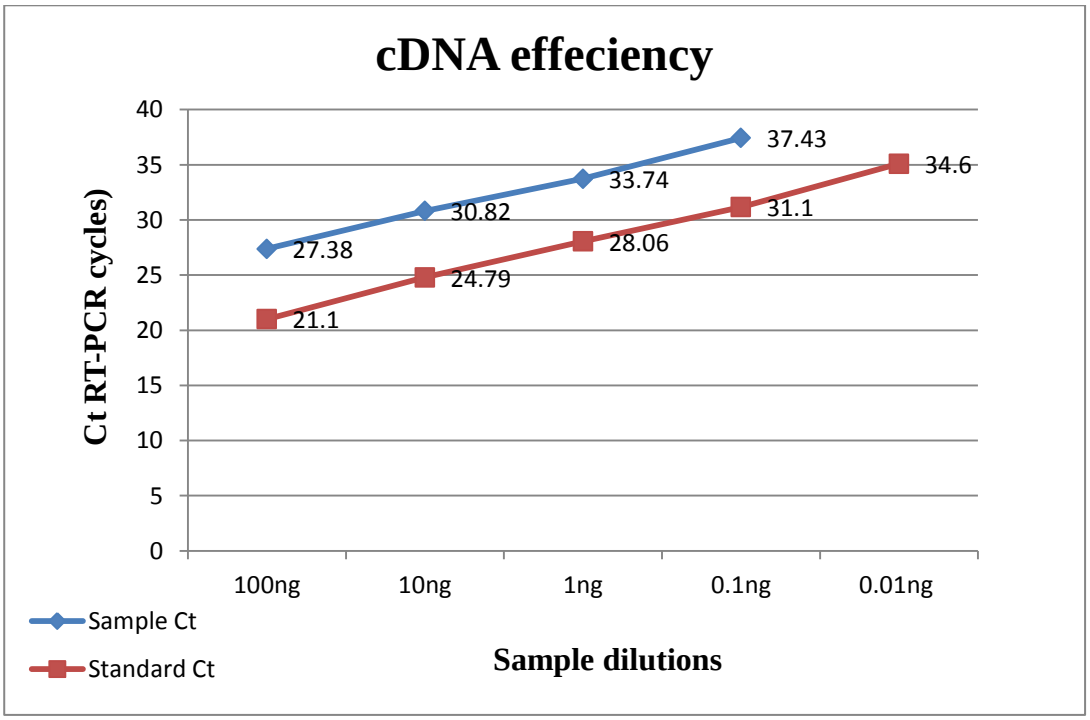

Figure 8. RT-PCR assay amplification efficiency of gene *SAR0216*

cDNA template and internal reference gene 16S rRNA, 100 ng were used to make 10 fold dilution series and using Ct slope method within 5 data points (concentrations) and 5 log dilution value, calculated efficiency for target gene assay was 98% and amplification factor 1.98.

Table 5. Primers/probes, genes base sequences and predicted sizes of qPCR products for *S. aureus* lipoprotein genes

Gene Symbol	Primers/probe	Oligonucleotide sequence (5'→3')	Nucleotide location within genome (bp)	Size of qPCR product (bp)
SAR0444	Forward	ATAACCCAGAAGCACCAATATAC	472145 to 472287	143 bp
	Reverse	GCCTTTTAGATTACCTGACCC		
	Probe	CCGACGCAAAAAGCGCCTAAA		
SAR0730	Forward	GAAATTAATCATCAGCATTATGGCG	766367 to 766454	88 bp
	Reverse	GATATCCTTTTCCAGAGTGGC		
	Probe	AACAGGTTGTGGTAAAAGCCAAGAG		
SAR0390	Forward	AGTGTTATCAATGTCAGCGG	426981 to 427147	167 bp
	Reverse	CAGCTTTTTTCACAGCATCTTC		
	Probe	TGCATGTGGCAATGATACTCCAAAAGA		
SAR0216	Forward	GCGGAATGTCCGGTA ATT TAG	251048 to 251189	142 bp
	Reverse	AGCCGCTCAGTACAACAAC		
	Probe	AATTGACTTGGCATCGAACTCTGCAAC		
SAR0340	Forward	GCAAAAGCGTTATATCCAAAAGTTC	385262 to 385379	118 bp
	Reverse	TCTCTTCTTTCATATCTGCAAGAC		
	Probe	AACGCTCTGAACCAGTTGCAGAA		
16S rRNA (endogenous control)	Forward	GGCAAGCGTTATCCGGAATT	514800 to 514900	111 bp
	Reverse	GTTTCCAATGACCCTCCACG		
	Probe	CCACGCGCGCTTTACGCCCA		

2.8 *Caenorhabditis elegans* as a model of bacterial infection

2.8.1 *C. elegans*, bacterial strains and growth conditions

Wild type *C. elegans* (Bristol N2), *E. coli* OP50 was cultured in Luria broth and incubated at 37°C overnight, nematode strain N2 and *E. coli* strain OP50 used in this work were obtained from Dr Darren Brooks, University of Salford.

2.8.2 Nematode Growth Medium (NGM) agar

The following were prepared in double distilled H_2O, 0.25% (w/v) Bacto Peptone (Sigma-Aldrich, UK), 0.3% (w/v) NaCl and 2% (w/v) Bacto Agar (Sigma-Aldrich, UK), autoclaved at 121°C for 30 min and cooled to 55°C. Media was supplemented with filter sterilized 5 µg/ml cholesterol in ethanol, 1 mM calcium chloride, 1 mM magnesium sulphate, 25 mM potassium phosphate buffer (pH 6) and 0.1% of nystatin (10,000 units /ml; Sigma-Aldrich, UK).

2.8.3 Nematode synchronization

It is important to synchronize nematode cultures in order to obtain large numbers of gravid adult animals within the same age timing for subsequent experiments of host infection. To synchronize *C. elegans* eggs via the hypochlorite method as described by Lewis and Fleming, 1995 (Lewis and Fleming, 1995), plates of a lawn of *E. coli* OP50 containing gravid hermaphrodites nematodes carrying many fertilized eggs visible in their uterus, plates were washed with 5 ml sterile ddH_2O into a 15 ml Falcon tube, 1 ml of 5% sodium hypochlorite and 500 µl of 5N NaOH were added and tubes were shaken vigorously for 30 s to lyse the worms and release their eggs which are protected from the sodium hypochlorite by their shells, tubes were shaken for 5 min to dissolve all adult worms. Tubes were then centrifuged at 1500 x g for 30 sec to pellet released eggs, supernatant was discarded until approx. 500 µl of worms and liquid remained, 14 ml of sterile ddH_2O were add to dilute the bleach, eggs were centrifuged at 1500 x g for 30s, supernatant was discarded again and wash process was repeated twice, after three washes, the pellet containing eggs was resuspend in 100 µl of ddH_2O and transferred to the edge of a clean NGM plate seeded with non-pathogenic strain *E. coli* OP50 lawn for ~ 52 h of growth at 25°C to use for killing assay.

2.8.4 Assay of *C. elegans* killing by *S. aureus*

C. elegans killing assay was carried out as previously reported (Garsin *et al.*, 2001, Sifri *et al.*, 2003), as follow. *S. aureus* plates were prepared as follows, *S. aureus* MRSA252 were incubated aerobically overnight with aeration at 37°C in tryptone soya broth supplemented with 5 µg/ml of nalidixic acid (Sigma-Aldrich, UK), 1:10 dilution of overnight culture was prepared in fresh media, 10 µl of diluted culture spread on 6-cm-diameter plates containing TS agar supplemented with 5 µg/ml of nalidixic acid. Inoculated plates were incubated at 37°C for 6 h and then allowed to equilibrate to room temperature for 30 min before being seeded with *C. elegans*. In each assay, between 30-40 L4 or young adult hermaphrodite nematodes were transferred from a lawn of *E. coli* OP50 to the prepared killing plates of *S. aureus* on the agar, just outside the lawn of bacteria and each assay experiment was repeated on five independent occasions and thus ~ 175 nematodes were followed. Plates were incubated at 20°C and examined at 24 h intervals under the dissecting microscope for viability, visually inspection of nematodes for movement and nematodes were considered dead when they failed to respond to gentle touch with a platinum wire pick. Nematodes that died as a result of getting stuck to the wall of the plate or been embedded under the agar were not included in the analysis data. *C. elegans* with food source *E. coli* OP50 incubated at 20°C and monitored every 12 h served as control equal numbers of synchronized worms were picked and fed on *E. coli* OP50 for 28 days, survival rate indicated that lifespan of control animals was 28 days. Nematode survival was plotted with the non-parametric analysis by Kaplan-Meier method using Minitab software, Kaplan-Meier estimate is the best method to measure the fraction of subjects living for a certain period of time after treatment (Bland and Altman, 1998).

2.8.5 Total RNA isolation

Total RNA was obtained from cells at late-log phase growth for the control samples and from each growth condition of *C. elegans* that had been infected with *S. aureus* MRSA252 using RNeasy kit (Qiagen), approx. 100 nematodes were transferred to 1.5 ml micro-centrifuge tube and volume made up to 1 ml with M9 buffer containing 1mM sodium azide to inhibit infected animals movement, nematodes were washed three times with M9 buffer and vortexed at 2000 ×g for 2 min, nematodes were lysed in buffer of (20 mM Tris-HCl pH 8 (Sigma-Aldrich, UK), 0.5% Triton X-100 (Fisher Scientific, UK), 0.5% Tween-20 (Bio-Rad), 2 mM EDTA (Sigma-Aldrich, UK) 10 mg/ml lysozyme, 1 mg/ml lysostaphin

and 1 mg/ml proteinase K (Qiagen) in 0.5 ml PCR tube, tubes were briefly centrifuged to bring nematodes to the bottom and incubated at 37°C for 30 min, 500 µl of RLT buffer were add to homogenize the lysate nematodes and centrifuged at 12000 ×g for 2 min at room temperature in homogenizer (Invitrogen, UK). RNA isolation was performed according to the manufacturer's instructions Qiagen RNeasy Mini Kit. Extracted RNA were treated with TURBO DNase (Ambion), 1 µl DNase (2 U) for up to 10 µg of RNA in a 50 µl reaction, DNase enzyme was inactivated by adding 0.1 volume of inactivation reagent and incubated for 5 min at room temperature. RNA was quantified by using highly sensitive and accurate fluorescence-based Qubit quantitation assays Qubit 2.0 Fluorometer (Invitrogen, UK). Total RNA was checked for DNA contamination via PCR, samples of high quality were used for cDNA library construction. Ribo-Zero Magnetic Gold Kit (Epidemiology), Low-Input (Epicentre Technologies) was used to remove unwanted 5S, 16S and 23S rRNAs from ~ 500 ng of starting material prior to sequencing analysis and creation of a sequencing library, rRNA depletion, quantity and quality assessment of each sample were carried out at Centre for Genomic Research (CGR)/ Liverpool University, UK. Final libraries were pooled in equimolar amounts using the Qubit and Bioanalyzer data, RNA was evaluated by Bioanalyzer and subsequently by qPCR using the Illumina Library Quantification Kit from Kapa (KK4854) on a Roche Light Cycler LC480II according to manufacturer's instructions. Ribo-Zero treated RNA samples were converted to cDNA libraries using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre Technologies) and sequenced on an Illumina HiSeq 2500 system, libraries were purified using AMPure XP beads, reads of each organism were mapped to the corresponding genome assembly to calculate expression level of each gene. A brief overview of *C. elegans* killing assay and RNA library preparation are shown in figure 9. The experiment aimed to detect differentially expressed transcripts in contrasting samples of three conditions: C16 (control at 16h), T16 (treated for 16h) and T40 (treated for 40h). Samples are a mixture of RNA collected from *S. aureus* MASA252 and *C. elegans* Bristol N2, sample names and associated conditions are listed in table 6.

Table 6. Details of RNA samples from three independent experiment of treated and control cells

Sample No	Group	Time point	Condition
1	Control	16h	Not treated
2	Control	16h	Not treated
3	Control	16h	Not treated
4	Test	16h	*S. aureus* treated
5	Test	16h	*S. aureus* treated
6	Test	16h	*S. aureus* treated
7	Test	40h	*S. aureus* treated
8	Test	40h	*S. aureus* treated
9	Test	40h	*S. aureus* treated

2.8.6 Alignment of reads to reference sequences

The reference genome used for alignment was combination of the genome references for *S. aureus* MASA252 and *C. elegans* Bristol N2, *S. aureus* MRSA252 genome reference sequences was obtained from the NCBI genome database, the sequence FASTA file was downloaded from: ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Bacteria/Staphylococcus_aureus_aureus_MRSA252_uid265/BX571856.fna. The *C. elegans* Bristol N2 genome reference sequences and annotations were downloaded from url: ftp://ftp.ensembl.org/pub/release-81/fasta/caenorhabditis_elegans/dna/Caenorhabditis_elegans.WBcel235.dna_sm.toplevel.fa.gz. Reads were aligned to the combined genome sequences using Tophat (version 2.1.0) (Kim *et al.*, 2013a), invoked the short read mapper Bowtie2 (version 2.2.5) (Langmead and Salzberg, 2012). The default mapping setting was used. Such an alignment setup informs the software of the read pair orientation and instructs it to report up to 1 alignment per read, if the number of available alignments is greater than 1, the alignment with the highest mapping quality is reported, when two or more alignments are mapped with equal quality, one of them is randomly selected and reported, read alignment statistics generated from samtools (Li, 2011).

2.8.7 Differential expression analysis

Gene expressions levels of individual transcripts were estimated from reads alignment files by using Cufflinks (Trapnell *et al.*, 2012), corresponding FPKM values were converted from count numbers and reported in this analysis. The number of fragment per kilobase per million base (FPKM) was used to show the gene expression levels and an FPKM value ≥1 was used as a threshold (P-value <0.05). The count data were split into two sub-sets: one containing genes belong to *S. aureus*, other containing genes for *C. elegans*. The two sub-data sets were analysed separately. The differential gene expression analyses were applied to the two sub-count data sets, respectively. The main processes of the analysis include data variation assessment, data modelling, model fitting, testing and the detection of differentially-expressed (DE) transcripts. DGE (Differential Gene Expression) analyses were performed in the R environment using the edgeR package (Robinson *et al.*, 2010). P-values associated with Fold Changec (FC) were adjusted for multiple testing using the False Discovery Rate (FDR) approach (Benjamini and Hochberg, 1995). Significantly differentially expressed genes were defined as those with FDR-adjusted P-value < 5%. Family-wise error rate (FWER) correction which controls the probability of committing of errors for any of tests, the concept of false discovery rate (FDR) allows to tolerate a certain number of tests to be incorrectly discovered by using threshold that can be declare tests as significant or not, all these analysis were carried out at Centre for Genomic Research (CGR)/ Liverpool University, UK.

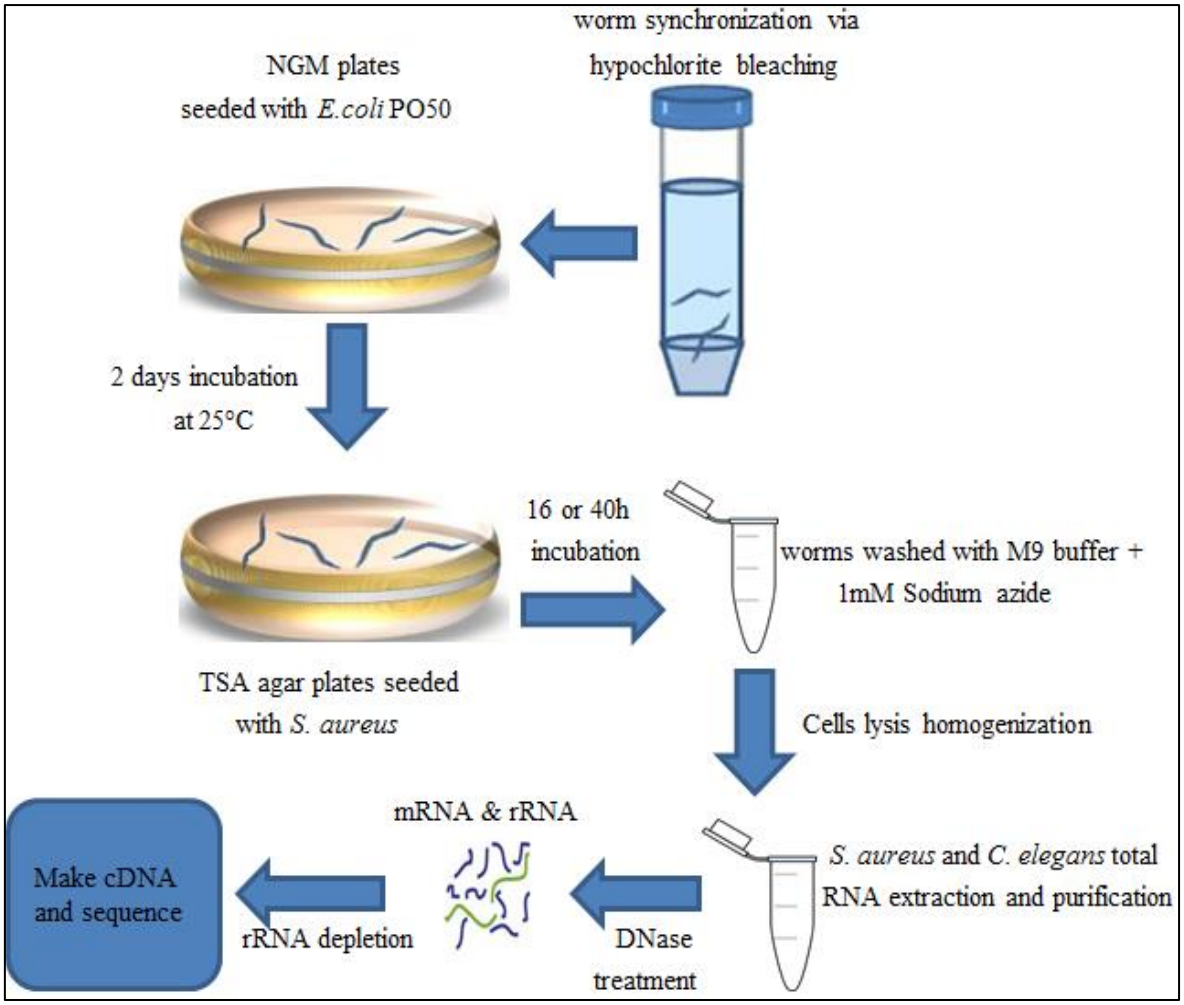

Figure 9. Schematic diagram illustrating an overview of *C. elegans* killing assay

S. aureus and total RNA isolation, starting from nematodes synchronization via hypochlorite bleaching and seeding nematodes in non-pathogenic *E.coli* OP50 for 48h before moving nematodes to plates seeded with pathogenic *S. aureus* MRSA252. Homogenized cells lyses were subjected to RNA extraction and followed by DNase treatment. rRNA was first removed from samples and cDNA libraries created using ScriptSeq Complete Gold Kit, Low-Input.

2.9 Differential expression analysis for *S. aureus* 8325-4 lipoprotein genes with RNA sequencing

All steps were performed by Dr Sari Al-Houfie

2.9.1 Bacterial culture

S. aureus 8325-4 was inoculated into 20 ml Iso-sensitest broth (Oxoid, Basingstoke, UK) and incubated with constant shaking at 37°C and 200 r.p.m. When the cells reached mid exponentially phase (OD 0.5) cells were harvested after 1, 12 and 24 h and their OD were 1.2, 6.7 and 10.4 respectively.

2.9.2 RNA extraction, RNA library preparation and sequencing

RNAprotect Bacteria Reagent and RNeasy Mini Kit (Qiagen) reagents were used to extract total RNA. DNase I (Qiagen) was included in the RNA extraction procedures to avoid DNA contamination. Purity of RNA was verified by NanoDrop spectrophotometer 2000 (Thermo Scientific) and accurate RNA concentration measurements were obtained via Qubit RNA Assay Kits with a Qubit 2.0 Fluorometer (Life Technologies, Paisley, UK). RNA-seq and data analysis were performed at the Centre for Genomic Research Centre for Genomic Research (CGR) at Liverpool University, UK. Ribosomal RNAs were depleted from the samples by RiboZero Magnetic kit Bacteria (Epicentre Technologies), 2 μg of starting material was used. RNA-seq libraries were prepared from 20-25 ng of the enriched material using Epicentre ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre Technologies) and the libraries were purified using AMPure XP beads after 13 cycles of amplification. The final libraries were pooled in equimolar amounts and quantity and quality of each pool was assessed by Bioanalyzer and subsequently by qPCR using the Illumina KAPA Library Quantification Kits (Kapa Biosystems, UK) on a Roche Light Cycler LC480II according to the manufacturer's instructions. All pools were sequenced on 3 lanes of the HiSeq 2000 using 2x100 bp paired-end sequencing with v3 chemistry. The mapping tools were Mapper Bowtie 2.1.0 using paired-end mapping mode. Analysis software was R version 3.0.1 and edgeR package version 3.0.4.

2.9.3 Data analysis

The results were based on analyses of three biological replicates, based on fold changes count, sample correlation analysis was performed to examine the variation within sample

replicates and between samples groups in order to evaluate differential expression of genes at the three time-points based on a cut off level of ≤5% for False Discovery Rate (FDR) values and 2-fold change in expression (FDR = ≤5% PLUS 2-fold change), the same criteria. Data for the 43 lipoproteins was extracted from this data.

Chapter three

Genetic analysis of *Staphylococcus aureus* lipoproteins

3. Genetic analysis of *Staphylococcus aureus* lipoproteins

3.1 Introduction

Staphylococcus aureus strains have shown host specialization and different phenotypic and genotypic characters, such as production of toxins, cell surface proteins and antibiotic resistance, may differ between populations of *S. aureus*. This genetic flexibility has led to the evolution of many virulent and drug-resistant strains (Holden *et al.*, 2004). Many complete *S. aureus* genome sequences are freely available to scientific researchers, these genomic sequence data have opened the doors to understand bacterial evolution and comparative genomics approaches to evaluate the relatedness and diversity of gene composition across different strains. Genetic diversity and population structure of *S. aureus* has been studied with different molecular techniques such as protein A gene (spa) typing (Frénay *et al.*, 1996), multilocus sequence typing (MLST) (Sakwinska *et al.*, 2009), pulse-field gel electrophoresis (PFGE) (David *et al.*, 2013) and whole-genome sequencing (WGS) which provide all sequence data required to predict *S. aureus* phenotype (Gordon *et al.*, 2014).

Sequencing of disease-causing epidemic methicillin-resistant *S. aureus* EMRSA-16 clone strain (MRSA252) with a genome size ~ 2.8 Mbp isolated from clinical cases has a clinical importance and globally prevalent lineage. MRSA252 strain has genetically diverse ~ 6% of their genome with a novel unique genetic components compared with the available sequenced strains at that time (Holden *et al.*, 2004). The simple approach to quantify genes relatedness is to compare certain DNA sequences for a group of genes, this method is relatively quick and inexpensive, while, it is very reproducible and available to almost any research group working in molecular centres. The availability of free databases of various bacterial genomes and computer software to compare sequences, allow comparative genomics techniques to play a role for evaluating relatedness and taxonomic studies in different aspects of molecular researches. Sequence of individual genes can provide some information about the differences between related strains within a given gene, such as random mutations and horizontal transfer. This study will show some comparative analyses of lipoprotein genes by using previously published lipoprotein genes of *S. aureus* strains and sequence of lipoprotein genes from two clinical isolate *S. aureus* with homologous lipoprotein genes of *S. aureus* MRSA252. One purpose of this study was to examine and compare the genetic features of lipoprotein genes in group of *S. aureus*

strains. More knowledge about the genotypic variation of *S. aureus* lipoprotein could help to investigate whether *S. aureus* lipoprotein of different strains genotypes are similar to each other, this statistical evidence for individual gene sequence divergence can help to reflect the general level of genome divergence. The goal of this experiment to determine how was the different *S. aureus* lipoprotein encoding genes are related to each other and whether the lipoprotein were common to all strains and whether some lipoproteins are specific to particular strains. Bioinformatics is also used to determine the random mutations within the genes. Moreover, single nucleotide polymorphism (SNP) analysis is used to compare relatedness and genetic variation of *S. aureus* lipoproteins.

Furthermore, understanding *S. aureus* phylogeny and in particular the virulence factors distribution is very important to understand the diversity of virulent lineages between strains. Phylogeny trees data have been generated for these strains allowing comparison of the sequenced lipoprotein data, comparison of the genes sequences was facilitated by using different software packages which enabled the visualization of generated data. Tajima's D is a widely used statistical genetics comparison test by comparing between the average number of pairwise differences and the number of segregating sites in a sample. This test distinguish between DNA sequence evolving randomly and one evolving under a non-random process, also to detect natural selection in nucleotide sequences (Tajima, 1989). The randomly evolving DNA sequence contains mutations that do not affect the function of the genes, these mutations called neutral mutations. However, the non-neutral mutations may causes cellular disorders and malfunction e.g. loss of pathogenicity (Li *et al.*, 2002).

3.2 Results

3.2.1 Detection of lipoprotein genes and its homologues in *S. aureus* strains

To examine the presence of predicted lipoprotein genes in *S. aureus* strains, three *S. aureus* strains (MRSA252, T1 and RN4282) were examined with PCR using designed primers for lipoprotein genes sequences taken from the NCBI database for *S. aureus* MRSA252 to amplify the target coding region of 50 lipoprotein genes in each strain under the same conditions. Examples of agarose gel electrophoresis results are shown in figures 10 and 11. The results of PCR work conducted are summarized in Table 7. PCR reactions were successful for most of the 50 tested lipoprotein genes which showed clear bands of the predicted sizes (e.g. see figures 10 and 11). Most genes were present in all strains and only a few genes were not detected in some strains. The genes were sequenced by Sanger sequencing technology to perform further BLAST multiple alignments and phylogenetic analysis. Four genes among the examined strains were variable in presence, *SAR1881* (putative lipoprotein) was not detected in strains RN4282 and T1, *SAR2104* (putative lipoprotein) was not found in RN4282, *SAR0953* (transport system extracellular binding lipoprotein) and *SAR1494* (putative lipoprotein) were not detected in RN4282 strain.

Standard BLAST tools were used for alignment of 50 lipoprotein genes sequence of *S. aureus* MRSA252 based against another 19 *S. aureus* strains were performed as an individual basis. The high alignment score gives an indication of high similarity between the aligned sequences, also the low E-value was generally close to zero, signifying the statistical significance of the pairwise alignment and reflecting the size of the database searched. Sequencing coverage results demonstrated that the analysis by BLAST search of the sequenced genes revealed that the nucleotide identities of 28 lipoprotein genes detected in all strains were well conserved in all 20 strains showing high similarity scores between them with at least 98% and 100% of sequence query coverage. A further 11 genes showed minor changes of nucleotide residues with a similarity ranging between 100-96% and the non-synonyms mutation were more frequent than the synonyms mutation. Further, among the examined genes 11 lipoprotein genes had strong nucleotides changes leading to large amino acid residues variations, any sequences pairwise alignment with E-value less than $1e^{-04}$ can be considered related with an error rate of less than 0.01%. BLAST alignments search results are presented in appendix tables 1-44.

SAR0442, *SAR0444* and *SAR0443* genes despite their quite a small size (771bp, 782bp and 699bp respectively) were shown to have very changeable sequences with a high rate of synonymous and non-synonymous substitutions sites, the total number of mutations within the strains MRSA252, T1 and RN4282 were 343, 190 and 272 sites respectively. These SNPs were present in a range of sites, as indicated in table 8. *SAR0439* gene was extremely variable among all strains and shown the highest number of mutations sites in most strains, the total mutation sites 354 with 126 synonymous sites and the remaining 98 SNPs were non-synonymous. *SAR0730* gene was relatively conserved in the amino acids structure between the strains except 4 strains (RF122, M013, ST398 and LGA251) were found to be slightly modified and showed considerable mutation sites compared to the other strains as shown in table 8. *SAR1189* had a different pattern of amino acid diversity, four strains (LGA251, RF122, ST398 and M013) had an altered truncated gene sequences and revealed strong modification in their structure, while other strains had very slight nucleotide changes. Although, *SAR0438* gene was found relatively to be variable between the examined strains sequences as lowest nucleotide similarities among the sequences were up to 83%, while, *SAR1558* gene was found in 5 out of 20 strains with similarity percentage between 96-100% (appendix 1, table 8 and 48). In the other hand, the small sized 357bp *SAR1288* gene was detected in 7 strains of all tested strains, but this gene was very unchanged and showed very minor amino acids variation (appendix 1, table 47). The smallest lipoprotein gene *SAR0706* was detected in 4 strains only but with no nucleotide alteration in any strain (appendix 1, table 46).

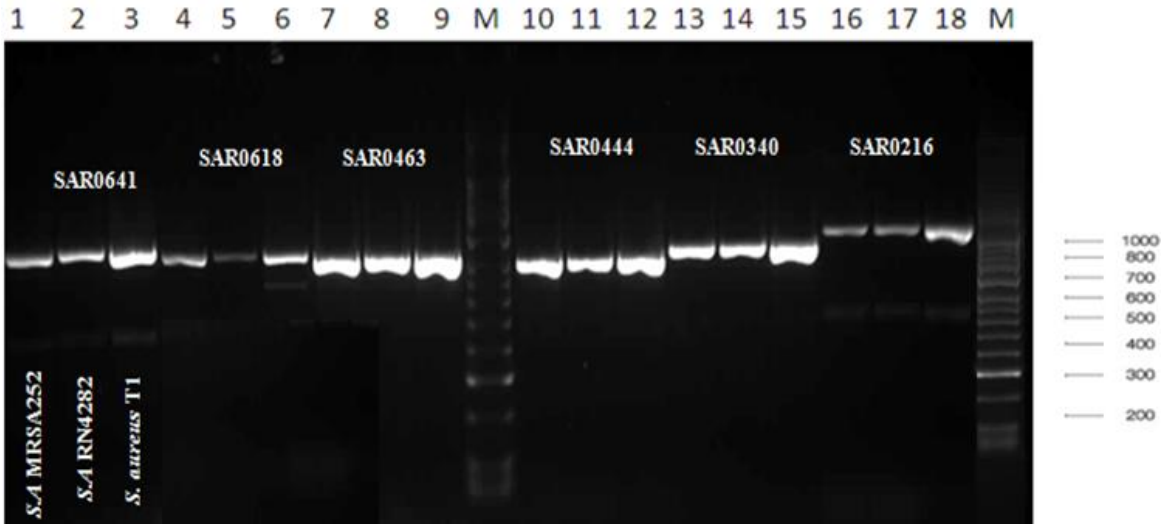

Figure 10. A 1.5% agarose gel image showing PCR products of 6 lipoprotein genes fragments

PCR products were stained with GelRed. (M) 50 bp Hyperladder II was used as molecular weight marker (Bioline), lane 1, 2, 3: *SAR0641* gene (930 bp), lane 4, 5, 6: *SAR0618* gene (888 bp), lane 7, 8, 9: *SAR0463* gene (843 bp), lane 10, 11, 12: *SAR0444* gene (795 bp), lane 13, 14, 15: *SAR0340* gene (855 bp), lane 16, 17, 18: *SAR0216* gene (969 bp), fragments for *S. aureus* strains (MRSA252, RN4282 and T1).

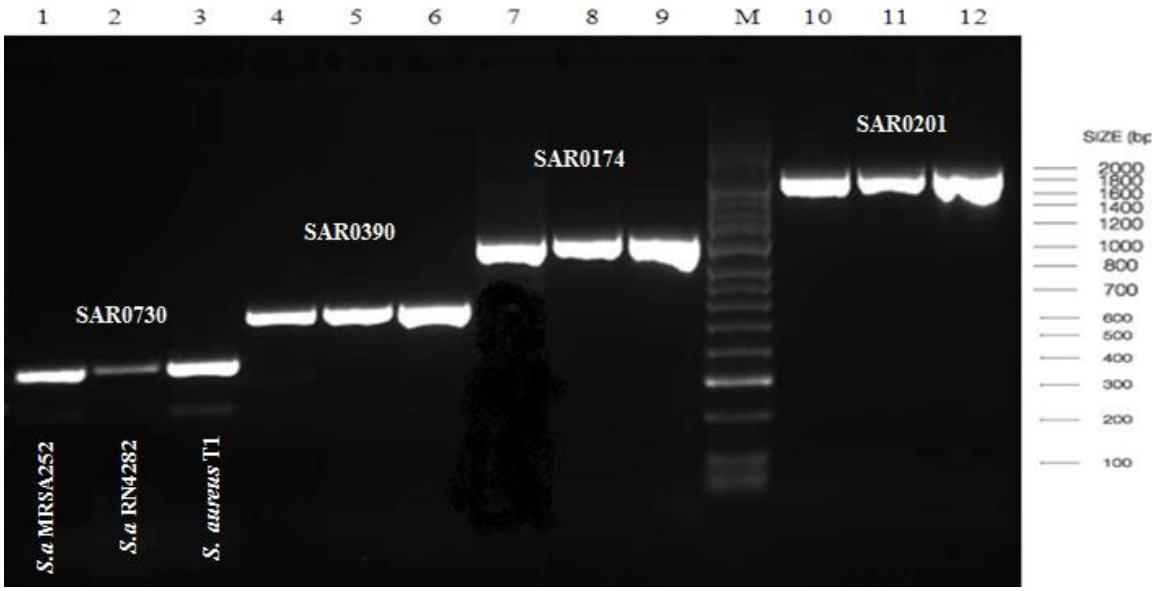

Figure 11. A 1.5% agarose gel image showing PCR products of 4 lipoprotein genes fragments

PCR Products were stained with GelRed. (M) 50 bp Hyperladder II was used as molecular weight marker (Bioline), lane 1, 2, 3: *SAR0730* gene (390 bp), lane 4, 5, 6: *SAR0390* gene (573 bp), lane 7, 8, 9: *SAR0174* gene (975 bp), lane 10, 11, 12: *SAR0201* gene (1776 bp), fragments for *S. aureus* strains (MRSA252, RN4282 and T1).

3.2.2 *S. aureus* lipoprotein nucleotide diversity (π), mutation rate and distribution

DNA variation information can be used to study genetic diversity within and between strains, and to conclude the population genetics parameters in statistic figures, also to investigate the mechanisms of nucleotide changes. The total 50 lipoprotein genes for 20 *S. aureus* strains were subjected to statistical calculations by DnaSP program for the analysis of nucleotide polymorphism from aligned DNA data, including the noncoding and synonymous or nonsynonymous sites. Genetic diversity analysis among examined lipoprotein genes for 20 *S. aureus* strains search has showed considerable high similarity between the majorities of these sequenced genes, data are summarized in table 8. The available information about these genes in UniProt database give a strong probability for a similar biological relationship between large number of these genes, however table 9 outlines some potential lipoprotein functions in *S. aureus* MRSA252 and function of similar genes in other *S. aureus* strains, these proteomics data were obtained from UniProt database (Consortium, 2014). BLAST searches for each lipoprotein gene gives a relative idea about their amino acid sequences, estimates of molecular variation and identifying individual gene sequence similarities with other appropriate lipoprotein of *S. aureus* strains.

As noted above, DNA analysis revealed variable sequence variation between *S. aureus* lipoprotein, BLAST alignment of 50 lipoprotein gene sequence of *S. aureus* MRSA252 against another 19 *S. aureus* strains were performed in an individual basis. There were no significant differences in gene sequence of 28/50 genes (56%) *SAR0794*, *SAR0641*, *SAR0618*, *SAR0463*, *SAR0390*, *SAR0340*, *SAR0216*, *SAR0201*, *SAR2496*, *SAR0872*, *SAR1106*, *SAR1608*, *SAR1995*, *SAR2179*, *SAR2368*, *SAR2499*, *SAR2500*, *SAR2536*, *SAR2546*, *SAR0794*, *SAR1034*, *SAR0463*, *SAR0118*, *SAR0206*, *SAR0230*, *SAR1495*, *SAR0396* and *SAR0872*) which were very highly conserved and have relative similarity score among them to show at least 98% nucleotide similarity and 100% query cover of all sequences, while the E-value was zero for all of these alignment genes, this low value present a great statistical significance of a given pairwise alignment. However, 11 genes had nucleotide similarity with a range of 94-97% with E-value was nil and up to 68% query cover of gene sequence. The other 11 genes were highly variable in the gene size and genetic structures, 5 genes were not detected in most strains (*SAR1558*, *SAR1288*, *SAR0706*, *SAR1831* and *SAR2104*), meanwhile 6 genes (*SAR0442*, *SAR0443*, *SAR1189*, *SAR0438*, *SAR0444* and *SAR0439*) exhibited strong genetic variation and showed an

extremely changeable structure coupled with low query cover and insignificant E-value for some truncated genes, these genes contained a large number of mutations compared with the other examined genes, BLAST alignment for all these genes are given in appendix A.

The number of mutations (SNPs) was highly variable between the examined genes, where the lowest mutation number was found in the relatively small size 363bp *SAR2500* gene with only 3 SNPs, however the highest number was 354 mutations within 737bp of *SAR0439* gene. In order to better understand of genetic variation in the form of single nucleotide polymorphisms (SNPs), nucleotide diversity was assessed by using DnaSP software package. The results (table 8) indicate the number of SNPs, mutations number and nucleotide diversity per lipoprotein gene for examined *S. aureus* strains unless the lipoprotein gene was not detected in the most strains, the calculations present the proportion of synonymous (silent; *ds*) and non-synonymous (amino acid-changing; *dn*) substitution rates, the mean pairwise percentage nucleotide diversity, number of mutations, invariable sites (monomorphic and polymorphic), singleton variable sites and parsimony informative sites (2, 3 or 4 nucleotide variants) of common 44 lipoprotein genes.

To measure the degree of polymorphism within the lipoprotein genes the average pairwise nucleotide diversity for the whole concatenated head-to-tail lipoprotein genes was 0.018 among all sequences, the highest rate of nucleotide diversity of candidate genes was detected in *SAR0439* gene with 0.1272, while the lowest rate was found within *SAR2500* gene 0.0008. Twenty three lipoprotein genes showed low nucleotide diversity with a value less than 0.010, these genes were fairly invariant between all strains. Furthermore BLAST searches revealed similar results with high similarity alignment for these genes. Gene alignments involved in this study are given in the appendix. Eleven lipoprotein genes were found to be highly variant in all strains to indicate extreme nucleotide diversity with value range 0.022 and 0.127, also BLAST analyses for these genes indicated a high level of DNA variation and large number of synonymous and non-synonymous mutations.
The results suggest that overall there were 50 different lipoproteins of which 44 were common to all examined strains, of these 28 genes had highly conserved sequences, 6 genes occurred in ~ 8 strains and 5 genes that only appeared in few tested strains, while 11 genes were highly variable in their size and genetic structures.

Table 7. List of 50 lipoprotein genes tested by PCR in three *S. aureus* strains (MRSA252, RN4282 and T1)

Genes not detected in some strains are indicated with (●). ORF indicates the gene locus in MRSA252 strain.

Gene ORF	MRSA252	RN4282	T1	Gene ORF	MRSA252	RN4282	T1
SAR0118	✓	✓	✓	*SAR1066*	✓	✓	✓
SAR0174	✓	●	✓	*SAR1106*	✓	✓	✓
SAR0201	✓	✓	✓	*SAR1189*	✓	✓	✓
SAR0206	✓	●	✓	*SAR1288*	✓	✓	✓
SAR0216	✓	✓	✓	*SAR1494*	✓	●	✓
SAR0230	✓	✓	✓	*SAR1495*	✓	✓	✓
SAR0340	✓	✓	✓	*SAR1558*	✓	✓	✓
SAR0390	✓	●	✓	*SAR1608*	✓	✓	✓
SAR0396	✓	✓	✓	*SAR1831*	✓	✓	✓
SAR0438	✓	✓	✓	*SAR1879*	✓	✓	✓
SAR0439	✓	✓	✓	*SAR1881*	✓	●	●
SAR0442	✓	✓	✓	*SAR1995*	✓	✓	✓
SAR0443	✓	✓	✓	*SAR2104*	✓	●	✓
SAR0444	✓	✓	✓	*SAR2179*	✓	✓	✓
SAR0463	✓	✓	✓	*SAR2268*	✓	✓	✓
SAR0618	✓	✓	✓	*SAR2363*	✓	✓	✓
SAR0641	✓	✓	✓	*SAR2368*	✓	✓	✓
SAR0706	✓	✓	✓	*SAR2457*	✓	✓	✓
SAR0730	✓	✓	✓	*SAR2496*	✓	✓	✓
SAR0790	✓	✓	✓	*SAR2499*	✓	✓	✓
SAR0794	✓	✓	✓	*SAR2500*	✓	✓	✓
SAR0872	✓	✓	✓	*SAR2504*	✓	✓	✓
SAR0953	✓	●	✓	*SAR2536*	✓	✓	✓
SAR1011	✓	✓	✓	*SAR2546*	✓	✓	✓
SAR1034	✓	●	●	*SAR2554*	✓	✓	✓

Table 8. Phylogenetic analysis of 44 lipoprotein genes for 20 *S. aureus* strains, calculated by DnaSP program
*Synonymous (*ds*), *non-synonymous (*dn*).

Gene ORF/ size (bp)	Variable sites (polymorphic)	Number of mutations	*ds* / *dn* mutation number	Singleton variable sites (one nucleotide variant)	Parsimony informative sites (2, 3 or 4 nucleotide variants)	Nucleotide diversity (π)
SAR0118/ 993	15	15	11/4	2 nucleotide variants: 4 3 nucleotide variants: 0	2 nucleotide variants: 11 3 nucleotide variants: 0	0.0040
SAR0174/ 975	54	57	34/23	2 nucleotide variants: 22 3 nucleotide variants: 0	2 nucleotide variants: 29 3 nucleotide variants: 3	0.0159
SAR0201/ 1776	186	189	148/38	2 nucleotide variants: 157 3 nucleotide variants: 0	2 nucleotide variants: 26 3 nucleotide variants: 3	0.0157
SAR0206/ 1272	30	31	25/6	2 nucleotide variants: 12 3 nucleotide variants: 0	2 nucleotide variants: 17 3 nucleotide variants: 1	0.0072
SAR0216/ 969	44	44	25/16	2 nucleotide variants: 22 3 nucleotide variants: 0	2 nucleotide variants: 22 3 nucleotide variants: 0	0.0117
SAR0230/ 1476	57	57	38/19	2 nucleotide variants: 23 3 nucleotide variants: 0	2 nucleotide variants: 34 3 nucleotide variants: 0	0.0096
SAR0340/ 855	23	23	16/7	2 nucleotide variants: 15 3 nucleotide variants: 0	2 nucleotide variants: 8 3 nucleotide variants: 0	0.0042
SAR0390/ 573	44	46	34/12	2 nucleotide variants: 0 3 nucleotide variants: 35	2 nucleotide variants: 7 3 nucleotide variants: 2	0.0114
SAR0396/ 627	15	16	12/4	2 nucleotide variants: 4 3 nucleotide variants: 0	2 nucleotide variants: 10 3 nucleotide variants: 1	0.0058
SAR0438/ 699	193	218	92/73	2 nucleotide variants: 42 3 nucleotide variants: 1	2 nucleotide variants: 129 3 nucleotide variants: 18	0.0783
SAR0439/ 737	291	354	126/98	2 nucleotide variants: 74 3 nucleotide variants: 5	2 nucleotide variants: 160 3 nucleotide variants: 46 4 nucleotide variants: 6	0.1272
SAR0442/ 771	286	343	129/103	2 nucleotide variants: 80 3 nucleotide variants: 0	2 nucleotide variants:153 3 nucleotide variants: 49 4 nucleotide variants: 4	0.1111
SAR0443/ 699	240	272	130/89	2 nucleotide variants: 123 3 nucleotide variants: 4	2 nucleotide variants: 87 3 nucleotide variants: 24 4 nucleotide variants: 2	0.0757

Table 8-continued

Gene ORF/ size (bp)	Variable sites (polymorphic)	Number of mutations	*ds* / *dn* mutation number	Singleton variable sites (1 or 2 nucleotide variant)	Parsimony informative sites (2, 3 or 4 nucleotide variants)	Nucleotide diversity
SAR0444/ 795	177	190	92/73	2 nucleotide variants: 25 3 nucleotide variants: 1	2 nucleotide variants: 139 3 nucleotide variants: 12	0.0791
SAR0463/ 843	50	51	41/10	2 nucleotide variants: 20 3 nucleotide variants: 0	2 nucleotide variants: 29 3 nucleotide variants: 1	0.0159
SAR0618/ 888	22	23	16/7	2 nucleotide variants: 7 3 nucleotide variants: 0	2 nucleotide variants: 14 3 nucleotide variants: 1	0.0061
SAR0641/ 930	8	8	8/0	2 nucleotide variants: 3 3 nucleotide variants: 0	2 nucleotide variants: 5 3 nucleotide variants: 0	0.0552
SAR0730/ 396	48	48	24/24	2 nucleotide variants: 3 nucleotide variants:	2 nucleotide variants: 31 3 nucleotide variants: 0	0.0252
SAR0790/ 1029	88	90	75/15	2 nucleotide variants: 18 3 nucleotide variants: 1	2 nucleotide variants: 68 3 nucleotide variants: 1	0.0208
SAR0794/ 879	12	12	7/5	2 nucleotide variants: 8 3 nucleotide variants: 0	2 nucleotide variants: 4 3 nucleotide variants: 0	0.0025
SAR0872/ 822	28	28	21/7	2 nucleotide variants: 13 3 nucleotide variants: 0	2 nucleotide variants: 15 3 nucleotide variants: 0	0.0083
SAR0953/ 1656	90	91	70/21	2 nucleotide variants: 29 3 nucleotide variants: 0	2 nucleotide variants: 60 3 nucleotide variants: 1	0.0146
SAR1011/ 960	59	59	44/15	2 nucleotide variants: 18 3 nucleotide variants: 0	2 nucleotide variants: 41 3 nucleotide variants: 0	0.0173
SAR1034/ 1101	9	9	7/2	2 nucleotide variants: 3 3 nucleotide variants: 0	2 nucleotide variants: 6 3 nucleotide variants: 0	0.0017
SAR1066/ 627	28	28	20/8	2 nucleotide variants: 3 3 nucleotide variants: 0	2 nucleotide variants: 25 3 nucleotide variants: 0	0.0141
SAR1106/ 879	16	16	12/4	2 nucleotide variants: 12 3 nucleotide variants: 0	2 nucleotide variants: 4 3 nucleotide variants: 0	0.0024
SAR1189/ 954	115	113	48/65	2 nucleotide variants: 4 3 nucleotide variants: 0	2 nucleotide variants: 110 3 nucleotide variants: 1	0.0675
SAR1494/ 909	69	74	48/26	2 nucleotide variants: 20 3 nucleotide variants: 1	2 nucleotide variants: 44 3 nucleotide variants: 4	0.0486
SAR1608/ 582	9	9	4/5	2 nucleotide variants: 6 3 nucleotide variants: 0	2 nucleotide variants: 3 3 nucleotide variants: 0	0.0020

Table 8-continued

ORF/ size (bp)	Variable sites (polymorphic)	Number of mutations	*ds* / *dn* mutation number	Singleton variable sites (1 or 2 nucleotide variant)	Parsimony informative sites (2, 3 or 4 nucleotide variants)	Nucleotide diversity
SAR1879/ 555	37	38	14/24	2 nucleotide variants: 5 3 nucleotide variants: 0	2 nucleotide variants: 31 3 nucleotide variants: 1	0.0222
SAR1881/ 846	21	21	36/31	2 nucleotide variants: 18 3 nucleotide variants: 0	2 nucleotide variants: 3 3 nucleotide variants: 0	0.0415
SAR1995/ 1200	24	24	18/6	2 nucleotide variants: 12 3 nucleotide variants: 0	2 nucleotide variants: 12 3 nucleotide variants: 0	0.0042
SAR2104/ 465	15	16	9/7	2 nucleotide variants: 15 3 nucleotide variants: 0	2 nucleotide variants: 1 3 nucleotide variants: 0	0.0104
SAR2179/ 873	18	18	15/3	2 nucleotide variants: 7 3 nucleotide variants: 0	2 nucleotide variants: 11 3 nucleotide variants: 0	0.0042
SAR2268/ 984	46	47	37/10	2 nucleotide variants: 6 3 nucleotide variants: 0	2 nucleotide variants: 39 3 nucleotide variants: 1	0.0137
SAR2363/ 783	55	57	57/22	2 nucleotide variants: 14 3 nucleotide variants: 0	2 nucleotide variants: 39 3 nucleotide variants: 2	0.0227
SAR2368/ 909	18	18	15/3	2 nucleotide variants: 8 3 nucleotide variants: 0	2 nucleotide variants: 10 3 nucleotide variants: 0	0.0042
SAR2457/ 630	50	52	24/15	2 nucleotide variants: 7 3 nucleotide variants: 0	2 nucleotide variants: 41 3 nucleotide variants: 2	0.0211
SAR2499/ 600	18	18	14/4	2 nucleotide variants: 10 3 nucleotide variants: 0	2 nucleotide variants: 8 3 nucleotide variants: 0	0.0064
SAR2500/ 363	3	3	1/2	2 nucleotide variants: 3 3 nucleotide variants: 0	2 nucleotide variants: 0 3 nucleotide variants: 0	0.0008
SAR2504/ 780	34	34	26/8	2 nucleotide variants: 5 3 nucleotide variants: 0	2 nucleotide variants: 29 3 nucleotide variants: 0	0.0115
SAR2536/ 942	44	46	34/9	2 nucleotide variants: 20 3 nucleotide variants: 1	2 nucleotide variants: 22 3 nucleotide variants: 1	0.0109
SAR2546/ 456	10	10	5/5	2 nucleotide variants: 5 3 nucleotide variants: 0	2 nucleotide variants: 5 3 nucleotide variants: 0	0.0051
SAR2554/ 1599	55	56	46/10	2 nucleotide variants: 14 3 nucleotide variants: 0	2 nucleotide variants: 40 3 nucleotide variants: 1	0.0093

Table 9. Comparison of lipoprotein function in *S. aureus* MRSA252 and other *S. aureus* strains according to the UniProt database

Gene Id	ORF	*S. aureus* MRSA252	Other *S. aureus*
gi|446967868	*SAR0118*	sirA lipoprotein	Iron compound ABC transporter iron compound-binding protein SirA, periplasmic binding protein
gi|49240544	*SAR0174*	Putative lipoprotein	Sulfonate/nitrate/taurine transport system substrate-binding protein
gi|49240571	*SAR0201*	RGD-containing lipoprotein	ABC transporter, substrate-binding protein, family 5
gi|49240576	*SAR0206*	Putative extracellular sugar-binding lipoprotein	Maltose/maltodextrin ABC superfamily ATP binding cassette transporter
gi|49240586	*SAR0216*	Putative lipoprotein	ABC transporter, solute-binding protein
gi|49240600	*SAR0230*	Putative extracellular solute-binding lipoprotein	Peptide ABC transporter, ABC transporter, substrate-binding protein
gi|49240702	*SAR0340*	Putative lipoprotein	Lipoprotein, hypothetical protein
gi|49240750	*SAR0390*	Putative lipoprotein	Putative lipoprotein, conserved hypothetical protein, propeptide
gi|49482628	*SAR0396*	Hypothetical protein	Putative lipoprotein, lipoprotein
gi|49482665	*SAR0438*	Lipoprotein	Zn-binding lipoprotein adcA, ribulose-phosphate 3-epimerase, Similar to Zn-binding lipoprotein adcA
gi|49240795	*SAR0439*	Putative lipoprotein	Tandem lipoprotein
gi|49240796	*SAR0442*	Putative membrane protein	Lipoprotein, tandem lipoprotein
gi|49240797	*SAR0443*	Putative lipoprotein	Lipoprotein.
gi|49240798	*SAR0444*	Putative lipoprotein	Tandem lipoprotein, conserved hypothetical protein
gi|49240817	*SAR0463*	Putative lipoprotein	Methionine ABC transporter substrate-binding protein, lactococcal lipoprotein
gi|49240968	*SAR0618*	Putative transport system lipoprotein	Vitamin B12 ABC transporter substrate-binding
gi|49482859	*SAR0641*	ABC transporter	Lipoprotein, hypothetical protein, iron-repressed lipoprotein
gi|49241045	*SAR0706*	Putative membrane protein	Hypothetical lipoprotein
gi|49240968	*SAR0618*	Putative transport system lipoprotein	Vitamin B12 ABC transporter substrate-binding
gi|49482859	*SAR0641*	ABC transporter	Lipoprotein, hypothetical protein, iron-repressed lipoprotein

Table 9-continued

Gene Id	ORF	*S. aureus* MRSA252	Other *S. aureus*
gi|49241045	*SAR0706*	Putative membrane protein	Hypothetical lipoprotein
gi|49241062	*SAR0730*	Putative lipoprotein	Hypothetical protein
gi|49482993	*SAR0790*	Lipoprotein	Iron (Fe+3) ABC superfamily ATP binding cassette transporter, iron complex transport system substrate-binding protein
gi|49241126	*SAR0794*	Putative lipoprotein	Conserved hypothetical protein
gi|49483071	*SAR0872*	Lipoprotein	ABC transporter substrate-binding protein, methionine ABC transporter substrate-binding protein
gi|49241280	*SAR0953*	Transport system extracellular binding lipoprotein	Ferrichrome ABC transporter lipoprotein,
gi|49241332	*SAR1011*	Transport system extracellular binding lipoprotein	Ferrichrome ABC transporter, putative iron transport protein
gi|49241354	*SAR1034*	Putative quinol oxidase polypeptide II precursor	Cytochrome aa3 quinol oxidase, subunit II
gi|49241384	*SAR1066*	Putative lipoprotein	Cell-wall binding lipoprotein
gi|49241424	*SAR1106*	Putative transport system extracellular binding lipoprotein	Putative iron transport lipoprotein SirF, heme uptake system protein IsdE
gi|49241505	*SAR1189*	Putative lipoprotein	Lipoprotein, hypothetical protein
gi|49483474	*SAR1288*	Lipoprotein	Hypothetical protein, putative lipoprotein
gi|49241801	*SAR1494*	Putative lipoprotein	Lipoprotein
gi|49241802	*SAR1495*	Putative lipoprotein	Lipoprotein, conserved hypothetical protein
gi|49483736	*SAR1558*	Lipoprotein	Putative lipoprotein
gi|49483780	*SAR1608*	Hypothetical protein	Lipoprotein
gi|49242122	*SAR1831*	Beta-lactamase precursor	Beta-lactamase
gi|49242169	*SAR1879*	Putative lipoprotein	Excalibur calcium-binding domain protein
gi|49242171	*SAR1881*	Putative lipoprotein	Lipoprotein, tandem lipoprotein
gi|49484144	*SAR1995*	Lipoprotein	Putative lipoprotein, hypothetical protein
gi|49242375	*SAR2104*	Putative lipoprotein	Lipoprotein, hypothetical protein

Table 9-continued

Gene Id	ORF	*S. aureus* MRSA252	Other *S. aureus*
gi|49242443	*SAR2179*	Putative membrane protein	Lipoprotein precursor, lipoprotein, membrane-embedded lipoprotein precursor
gi|49242526	*SAR2268*	Putative transport system binding lipoprotein	Iron citrate ABC transporter substrate-binding protein, putative ferrichrome-binding protein FhuD
gi|49484494	*SAR2363*	Molybdate-binding lipoprotein	Molybdate ABC superfamily ATP binding cassette transporter, binding protein, putative molybdate-binding protein
gi|49484499	*SAR2368*	Ferrichrome-binding lipoprotein precursor	Ferrichrome ABC transporter substrate-binding protein
gi|49484586	*SAR2457*	Hypothetical protein	Lipoprotein, conserved hypothetical protein
gi|49242751	*SAR2496*	Putative solute binding lipoprotein	Zn-binding lipoprotein adcA-protein, ribulose-phosphate 3-epimerase
gi|49242754	*SAR2499*	Putative lipoprotein	DSBA-like thioredoxine domain protein, lipoprotein
gi|49484626	*SAR2500*	Lipoprotein	Hypothetical protein
gi|49242759	*SAR2504*	Extracellular solute-binding protein	Extracellular solute-binding protein, family 3, amino acid ABC superfamily ATP binding cassette transporter
gi|49242792	*SAR2536*	Putative glycine betaine/carnitine/choline-binding lipoprotein precursor	Osmoprotectant transport system substrate-binding protein
gi|49242802	*SAR2546*	Putative lipoprotein	Lipoprotein, PF06998 family protein
gi|49242810	*SAR2554*	Oligopeptide transporter putative substrate binding domain	Nickel ABC transporter, nickel/metallophore periplasmic binding protein

3.2.3 Prediction of promoter sequences and promoters structure

The promoter sequence of the *S. aureus* MRSA252 lipoprotein genes are very little information are available. The promoter sequences for lipoprotein genes were determined by mapping the transcription start site for each gene and compared to *E. coli* promoter consensus -10 and -35 sequences (Hertz and Stormo, 1995). Figure 12, table 10 and 11 show the location of lipoprotein genes on the reference circular genome of *S. aureus* MRSA252 and labeled according to their position of the strands. Promoter sequences of *S. aureus* MRSA252 lipoproteins, upstream sequences for 100 nucleotides from start codon obtained from www.kegg.jp (table 11). Including the coding sequences of fur binding sites, the promoter sequences and 4 binding sites were predicted using the BPROM web-based software (Solovyev and Salamov, 2011). Operons were predicted using Genome 2D (genome2D.molgenrig.nl) and Artemis (http://www.sanger.ac.uk/science/tools/artemis) (Rutherford *et al*., 2000).

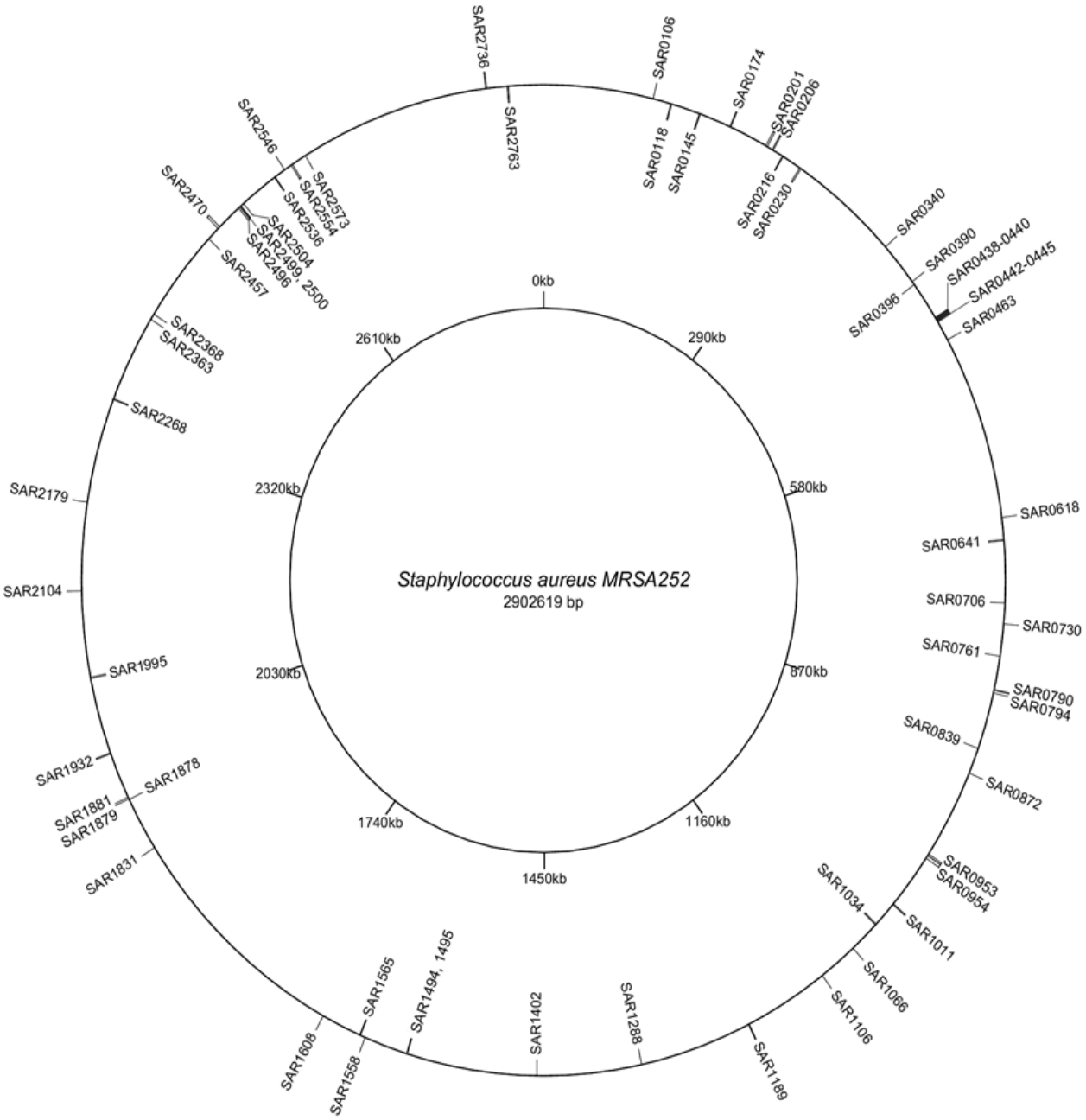

Figure 12. The distribution and location of lipoprotein genes mapped on the reference circular genome of *S. aureus* MRSA252 and labeled according to their position of the strands.

Table 10. An overview of genes encoding *S. aureus* lipoprotein and their regulating operons description, the promoter sequences and 4 binding sites were predicted using the BPROM web-based software (Solovyev and Salamov, 2011). Operons were predicted using Genome 2D (genome2D.molgenrig.nl) and Artemis (http://www.sanger.ac.uk/science/tools/artemis) (Rutherford *et al*., 2000).

ORF	Start	End	Strand	Gene	Probable operon structure
SAR0106	110754	111524	+		SAR106-SAR0107
SAR0118	128926	129918	-	*SirA*	SAR0116 (Sir B)-SAR0117 (SirC)-SAR0118 (SirA) Iron regulated
SAR0145	158898	159854	-		Monocistronic
SAR0174	191762	192736	+		SAR0173-SAR0174-SAR0175-SAR0176
SAR0201	231906	233681	+	*rlp*	SAR0199-SAR0200-SAR0201-SAR0202
SAR0206	239168	240439	+	*MalB*	SAR0205 (Mal A)-SAR0206 (MalA)-SAR0207 (MalC)-SAR0208 (MalD) Maltose transporter
SAR0216	251020	251988	-		SAR0214-SAR0215-SAR0216 sensor receptor/kinase
SAR0230	271601	273076	-		Substrate binding
SAR0340	385032	385886	+		SAR0340-SAR0341-SAR0342 Fe uptake
SAR0390	426962	427534	+		Monocistronic
SAR0396	430210	430836	-		SAR0396-SAR0397-SAR0398 (AhpF)-SAR0399 (AhpfC)
SAR0438	467510	468295	+		
SAR0439	468343	469116	+		SAR0448-SAR0449-SAR0440-
SAR0440	469147	469949	+		SAR0441-SAR0442-SAR0443-
SAR0442	469987	470757	+		SAR0444
SAR0443	470789	471589	+		SAR0440 Frameshift after codon 136
SAR0444	471608	472402	+		
SAR0445	472568	473389	+		SAR0445-SAR0446- SAR0447-SAR0448
SAR0463	491739	492581	+		SAR0462-SAR0463
SAR0618	666806	667693	+		Zinc, haem, cobalamin transporter
SAR0641	687790	688719	-	*MntC*	Fe, Zn, Cu transport
SAR0706	746977	747081	-		34 amino acids one transmembrane region
SAR0730	766363	766752	+		SAR0730-SAR0731-SAR0732
SAR0761	796705	797145	-		
SAR0790	828218	829246	+	*sstD*	SAR0787 (sstA)-SAR0788 (sstB)-SAR0789 (sstC)-SAR0790 (sstD) Fe transporter
SAR0794	831583	832461	+		Monocistronic
SAR0839	885498	886186	-		Monocistronic
SAR0872	910328	911149	+		Sars007-SAR0870-SAR0871-SAR0872

Table 10-continued

ORF	Start	End	Strand	Gene	Probable operon structure
SAR0953	995948	997603	+	*OppA*	SAR0949 (OppB)-SAR0950 (OppC)-SAR0951(OppD)-SAR0952 (OppF)-SAR0953 Oligopeptide transporter
SAR0954	997815	1000604	+		Monocistronic
SAR1011	1054214	1055173	+		Monocistronic Fe transport
SAR1034	1079932	1081032	-	*QoxA*	SAR1031 (Qox D)-SAR1032 (QoxC)-SAR1033 (QoxB)-SAR1034 (QoxA) Quinol oxidase
SAR1066	1112113	1112739	+		Monocistronic
SAR1106	1152065	1152943	+	*IsdE*	SAR1104 (IsdC)-SAR1105 (IsdD)-SAR1106 (IsdE)-SAR1107 (IsdF)-SAR1108 (SrtB)-SAR1109
SAR1189	1238170	1239123	+		Monocistronic
SAR1288	1352458	1352814	-		SAR1287-SAR1288-SAR1289-SAR1290 Fe regulated
SAR1402	1458802	1458062	-		Monocistronic
SAR1494	1589460	1590368	-		SAR1494-SAR1495
SAR1495	1590426	1591331	-		
SAR1558	1634418	1634852	+		SAR1557-SAR1558-SAR1559-SAR1560
SAR1565	1639484	1640417	-		Monocistronic
SAR1608	1681001	1681582	+		SAR1607-SAR1608
SAR1831	1913827	1914672	+	*BlaZ*	Monocistronic
SAR1878	1965275	1965757	-		Monocistronic
SAR1879	1966128	1966682	+		SAR1879-SAR1780
SAR1881	1967834	1968460	+		SAR1881-SAR1882-SAR1883
SAR1932	2011797	2012759	+		Monocistronic
SAR1995	2084777	2085976	-		SAR1995-SAR1996-SAR1997-SAR1998
SAR2104	2166873	2167337	+		SAR2103-SAR2104-SAR2105
SAR2179	2249989	2250861	+		SAR2179-SAR2180-SAR2181-SAR2182-SAR2183-sars023
SAR2268	2349338	2350321	-	*FhuD*	SAR2266-SAR2267-SAR2268 Mo transport
SAR2363	2433026	2433808	-	*ModA*	SAR2361 (ModC)-SAR2362 (ModB)-SAR2363 (ModA)
SAR2368	2439723	2438831	-		Monocistronic
SAR2457	2528101	2528730	-		Monocistronic
SAR2470	2541365	2543521	+		Monocistronic
SAR2496	2569326	2570873	-		Monocistronic Zinc binding/transport
SAR2499	2571954	2572553	-		SAR2499-SAR2500
SAR2500	2572572	2572934	-		

Table 10-continued

ORF	Start	End	Strand	Gene	Probable operon structure
SAR2504	2575960	2576739	-		SAR2502-SAR2503-SAR2504-SAR2505 Transporter
SAR2536	2615364	2616305	-	*OpucC*	SAR2535 (OpucD)-SAR2536 (OpucC)-SAR2537 (OpucB), SAR2538 (OpucA)
SAR2546	2627660	2628115	+		Monocistronic
SAR2554	2635602	2637200	-		SAR2549-(Opp1F)-SAR2550 (Opp1C)-SAR2551 (Opp1D)-SAR2552 (Opp-1B)-SAR2553 (Opp1A)-SAR2554
SAR2573	2652402	2653178	-		Monocistronic
SAR2736	2844639	2845582	+		SAR2736-SAR2738-SAR2739
SAR2763	2866243	2867367	-		SAR2763-SAR2764-SAR2765-SAR2766-SAR2767-SAR2768

Table 11. Promoter sequences of *S. aureus* MRSA252 lipoproteins

(Bold ORF gene promoters shown for initial genes in operon marked not lipoprotein), upstream sequences for 100 nucleotides from start codon obtained from www.kegg.jp. Key: Coding sequences in blue and underlined, -10 sequences highlighted in yellow, -35 sequences in red, fur binding site in blue. Promoter sequences and 4 binding sites were predicted using BPROM (Solovyev and Salamov, 2011).

ORF	Promoter sequences
SAR0106	TGTGGAATTTTTGATAATATTAAAGTGAAAAAGTGTTATAAATTGATATTAATATGTAATTTGCAAAAGTAAATCATTTTAAAAAGAAGAGAGTTGTAAGATG
SAR0118	AAATCATAAATATTTTATTGACTTAGGAAAAAATTTAATTCATACTAAACCGTGATAATGATTCTCATTGTCATACATCACGAAGGAGGCTAATTAGTCAATG
SAR0118	AAATCATAAATATTTTATTGACTTAGGAAAAAATTTAATTCATACTAAACCGTGATAATGATTCTCATTGTCATACATCACGAAGGAGGCTAATTAGTCAATG
SAR0145	TTTATAAAATACTTTTTAGCATTTAAATGTGGTACGCTATAAGTGTAATTTCATTGCATACATATTACACGATTAAGAATGTGAAGGGGACAGTTATCAAATG
SAR0173	TAGATATAACAATTCACGATTTAAGGGCTGTGTTTGGCATAGCCCTTTAGATATACATTTAATTCTTATTAAAATGGTAGGGATTAAAAGGGGGCTTGTCATG
SAR0174	GATAGCCACCTACTTAAGATTCGTAATGAAATTATGGAAACATTTGCATTGAATCATCATCAAGTTGAACCTGAATATTATTTATAAGGAGTGAGTGACGATG
SAR0201	GGTATTTGTATTATGTTCGTCGTTTTAGCATTTAATTTTATAGGTGATGCAGTGCGTGATGCGCTAGATCCAAGAATTCATTAAAAAGGTAGGGATAGATGTG
SAR0206	AGTGATGGCGAACGACAAGATTACACTAGCATTTGATATGAATAAGTGTCACTTTTTTGATGAAAAAACAGGAAATCGTATCGTCTAAGGGGGAGTATTCATG
SAR0216	ACAAGTCATCTATAAACATTTCTAAATATTTAACATTACTTATGCGTCATTTATTGCTAAAATTATTGTATTAAAATATACATAGAATTGATGGGATATCATG
SAR0230	TAATTACTGAACTTGAACAACATTTTACTTCAACAAAATAAAGTTTTAAAATATTAATTGTTGGGTTTTACATAGTAATTAGAAAGAGAGTGTTAATGAAATG
SAR0340	TTTAAGAGCCACGTATTATCTTGGCATTTAGGGAGTGTGAGATAGTACGAATGAGAGCATGTCTATGGGGATAATAGAATTTCTATAATGAGGTGTCAAAATG
SAR0390	GTAATAATTATGAATAGAATTAAAAACAAGGGGTAATACAATCTATATAGCATATAAGCTTTTGTTATGAGTTTCAAAAAAAGGAAGAGAGAGTGATATTATG
SAR0396	GCAAGGTCATCGTGACATTCCAAACTAATTCAATGTTACTCTATTAAACATGAATACATATTTTTTCAATTAAAGATAAACTAAGGAGTTATATTACATTATG
SAR0438	TATATGTATATTTTGGTAACGTAAAAGAGAAATATACAAAATAATTAATTATTTATATGAAAAGAGAATATAAATGAAGTATAAAACAGAGAGACGTGAAATG
SAR0439	AGACAACATTAAGAGATTATTTAAATTTTGGGCCGTCTGATGAAGATAGCTAAAGTTTATTAGATGTAGAAATAATTCTAAAACAGAGAGACGTGAAACGATG
SAR0442	GAAACAGCTGTAAGAGATTTTGTGAATTTTGGACCATCTGATGGAGGAGCTAAAGTTTATTAGGTGTAGAAATAATTCTAAAACAGAGAGATGTGAAAAGATG
SAR0443	GACTTTTCAAAAGAAGATAGGGATATTTCAGTCATTGATTATTTAAGTTATAAGCCAGCGAAAAAATAGTGTTGATAATTAAATATTAGGGTGTGAAATGATG
SAR0444	TTCAAAAGGTGAAAAAAGCGATTTGTCAGTAATAGATTCTTTGAATTTCCAGCCGGCGAAGGTGAATGAAGATGATGAATGATGAGGATGGTGTGTAACAATG
SAR0445	AGTTAGTTTTGTTTGATAACATAAAAGTGTAATAATCACTATTTTATTAAGTTGTATGAAATATTCAAATGTGTATAAAAATAAAGCAGAGAGATGTGAAATG
SAR0463	CGTTATTGTCCAAGTGATTCAAACGCTAGGGAATGTCCTAGCTAGATTCATACGTAGACATTGATGATATATAGTGAAGATTTTGAAAGGAATTGATAGAATG
SAR0618	AAAGTGGGATTAAAGGAACGCAGTTGGATGTAACCGCACAACTGCATAAAAGCCTCTAATCACTAAAGTTCAAGAGGCTTTAAAAAGGAGATGTGATAATGTG
SAR0641	TTCCTAGTGGTGCAACAATCGTACTTTGTACGTTTGTAATTTATATCATCACATTATTTTTCACAAAATTTACGAATAGAAAGAAACGAGGAAGTTTAATCATG
SAR0706	AACAACATTTATTAATATTCATGTTATTTCATGGAAGAAGAGATTTGTTATAATGTGAAGGTCGTCTATCTCTCAGACGTCAATTCGGCGTAGAGAGGAGGTG
SAR0730	GAGTCATGTAAAAATGTATAAAGATATTGATTAGAATTAGGATTATGTTGCTAATTCATGTTAAAATTAAAAAAGATTTAATGACGTTAAGGAGTTTTATATG

Table 11-continued

ORF	Promoter sequences
SAR0790	GAAAATTAATTTAATCACGATAAAGTCTGGAATACTATAACATAATTCATTTTCATAATAAACATGTTTTTGTATAATGAATCTGTTAAGGAGTGCAATCATG
SAR0794	TTTAAAGGGGTGATGCGCCAATTAAAGAAGGGTTAACAAGTATTTATTCGTATTTCACATCAAGCACACAGATTAAGCCAAAAGAGGAGAATATTATATTATG
SAR0872	TATCAACAGTTTTTATTTTAATTATTGTATTTATAATCCAATTCATTGGGGATTGGCTTACAAATAAACTTGATAAACGATAAATTGGGGGTTTCATTTTATG
SAR0953	AATAAGACCACAGCATTTTGTCTTTAGTACTGAAGAAGAAGCGGCACGACTACGAAAAAATTTGTTGGTAACACAAGATTAAGGGGAAGGGAGAAATGAAATG
SAR1066	TATACTTAGTAATGAAGATGTGTAATGTAATTGTTTAAAATTGATTTCCAAGCAGATTTTATTTATCATTTAATTTAAATAGCAAGTGGAGGTACAAGTAATG
SAR1104	ATAATAAATTGATTATAAAATAAATATTGACAATGATAATCATTATTATTTATGATTTTAGTAAAGACTTAAAAGCAATCATAAAAAGGAGGATTATGTTTTG
SAR1106	ATCATAAAAAGCAATATATGTTATTTGCAGTTGGCATTGTGTTAGCAACCATTTTACTTATTTCGGCACATTTATACAGCAGAAAGAGAGGTAACCAAGTTTG
SAR1189	AAAATATTGATGGTTAATGTAAAATATCAATATAAATAAGTTTTTTAAATTTATGTATGTTTATTTGATTCAAACAAAATAACTTAAGAGGAGAAAGTTTATG
SAR1288	ATTGATAATTTAAAACGTTTCAATTTAAAACCTCTAGGACGTTAGTTTGTTTCAAGACGATAGCAATGATATAATTTACGTAAAAAGGAGTGAATTTATCATG
SAR1402	ACATTTACAGTTTCTAAAGATGAGGTTAATAATTCGAGGTTAAGATAAAGATGTAATCAATACAAATACTATTTGTTGTTCATACAGGGAGGATATTTCAATG
SAR1494	TAATTACAGTGATAATGAGTCTATAAGGATTGAGGTCGAATGACTTGAAATCAGCTAATTTCTCTATATTCTAAACAAACACATAAATGGGTGATGAGCTATG
SAR1495	CAGTATACTAATTATTTATTAAAAGCTACTTGTTTTCATTGAGAATAAGTAGCTTTTTCAAACATAAAAGTTTTACAAACACATAAATAGGTGATGAAATATG
SAR1558	AGATATATGGTATTGGAACTCACTATGGCGAGTATTCGATCACATTTGAGCCGTTGAGAGTTTTTAAATATAAGGAAATATAAACAAAGGAGAGATACATATG
SAR1608	TTACCTCCTAGTGAAACGTTTGTGCATTTCAACTTTATACATTAAAATAATATCATAATAAGGATAAAAAATAATAGATATTGATTTTAGGGAGATAGTAATG
SAR1831	ACCTTCTTTCAAATATTTATAATAAACTATTGACACCGATATTACAATTGTAATATTATTGATTTATAAAAATTACAACTGTAATATCGGAGGGTTTATTTTG
SAR1878	TATTACTGCTTAAAATGACACGAACATGACAATATCATTTAATAAAATAATAAATTTAATTTGATATAATATAGTGTAATTATTTTTAGAGGTGTTGCTTTTG
SAR1879	AATCATGAAAATAACATATAAATATAGAGGAGATTTACCTTTGAATACAGAGAACAACAAGAATCAAAACCAATCTGCTAAAAATTCTGAAAGACGTGGCATG
SAR1881	CATGTACGACGCTTCAACAGAATAAAATTTCAGGGTAGCTCGGCTACCCTTCTTTTACGGAAAAATTAATTATACATAATCAAATCAAGGAGATAAAAAAATG
SAR1995	TTCAAAATTAACAAAAGCACAAAGTTTAGGTATTGAAATTTGGACAGAGCAACAATTTGTAGATAAGCAAAATGAATTAAATAGTTAGAGGGGTATGTCGATG
SAR2104	ATTTTACTAAACAAAAAAACGCCTACTAGTGTGAAAACGTATTGATTAATAGCGCCTATATGGAGTTTTAATATAAAAAGTAAGCAAAGGAGAAATGAGAATG
SAR2179	TTTTTTAATAATTAGTTAAATTTTTCCACTTATTGTTTCATACGTACTCATTAATGTGATAGAATAAATCCTATGTGAATATAATGATAGAGGTGGAAAAATG
SAR2268	CATACATTTTAGTAATATAAAAAGATTGAACGCTACTTGACAATGATAATTGTTATCAATAAAATAATAAATGAAGTTATACATATTAAGGAGTGGAACGATG
SAR2363	ATTATTTGAATCATTATTTAAAGAGAAAACGCTTCTTTTCAAAAATATTGTTAAGTGTATAATGAGAGTGTAAATATTTAATAATTGGGGGTATTTCATAATG
SAR2368	TGCGAAAATAAAAAAATAAATAACACTAGCTATTGTAAATGTATATGCAATCTTGTATAATTTTGGAATAATTGATAATCATTTTCAATAGGAGGAAATTATG
SAR2457	TATTGTTGACGTCTTTGGTAACAATCCATGAAATGATACACATGGTAGATATGTATTTCAAGCTTATTCAATGAATATCGAATTATAGGAGGAGATATGTATG
SAR2496	ATGATTCGACTTTATATTATTTATAAAAACAGTGCTTGTATTTTAAGTAGCATGAGTATATAATTTTAAAACGTAAAGATTACTATTTAGGAGGGTGACTATG
SAR2499	ATCATATAAGCAAGATTTTGATTCGAGACGATATTATCAACAACATGATGCGGATTATCATGAAGAAAATATGACGAACTAGATATGAACAGGAGTTATTATG
SAR2500	TTAAATGTATTGGTAGTAGTTGTAGCGCATTGATGTTTGACGGTGTATAGTATTAATTATATTGAAAGTTAGTTGGAAGTTGATAGTAGGAGTGGGAGCATTG

Table 11-continued

No	ORF	Promoter sequences
55	*SAR2504*	AACGCTTTTTAGCAAAAGTTCAACATATTGACTTATCTGGCGATTCAGATTAAAATATTTTTATTCCGATTAGAATAATAAGAATAAGGAGATATATTCTATG
56	*SAR2536*	ACTGCACTAGCATTAGGTGTTGATGCCTTATTAGCTTTAGTTGAAAAATGGGTAGTTCCCAAAGGCTTAAAAGTATCTGGATAATTAGGAGGCTAAGATAATG
57	*SAR2546*	TACAAAAAAATTATCAAAATTTAAATTTTTGAATAACATCATTTTTTAGTAAATGTTATAATTTCTTTATAATACAAATATCGAAAGAAGGCGACTGACTTTG
58	*SAR2554*	TGTTTGACATTTACATAAAAATAAGCAAATAATTGAGAAAAATAATCATTACGATTTGATTAAGTAATGCAACTTATCAATTTAGAAAGAGGAAAAGCAAATG
59	*SAR2573*	ATTAAAAGCCACGAGGATTTAATCATAAAATTTCCTCTAGTATTATTAATATTGGAATTTGAACTTAACTTTGCATCAAAAATAAAAATGGGGATGTGAAATG
60	*SAR2736*	TCAAGATTAGCTTATAATAATATTGAATAACATTATTTTAGTTTTTGTATAATTTGCCTATCAACTGAAAAACAGCTAATACAAAGGAGCTAATGAAGATATG
61	*SAR2763*	TAAAAAATAAAATTAAAATCCCAACAGTTTCTTAAAAAGTTATGAGCCACATAATAAAAATGTTATAATTGTGCATAATAAAACATAGGGAGAATGTTTAATG

3.2.4 Lipobox sequences

S. aureus lipoprotein lipobox features of 50 lipoprotein genes were examined and were found to be highly conserved and showing relatively constant amino acid residues frequency among all strains with very low observed changes. The signal sequence divided into three regions: n-region, h-region and c-region. The n-region is characterized by presence of the positive amino acids lysine and/or arginine, the h-region consists of hydrophobic amino acids, and the c-region has a characteristic region of four amino acids around the cleavage site that is very well conserved, a so-called lipobox. The search criteria for lipobox in this task was as following: methionine at position 1, followed by 2 to 6 amino acids with at least one positive charge lysine or arginine, followed by a hydrophobic stretch of 7 to 15 amino acids followed by the lipobox with 4 amino acids. Analysis of lipobox amino acids frequency of an approximately 950 distinct lipoproteins in 20 *S. aureus* strains revealed four distinct amino acids sequences at the C-terminal end of the signal peptide including the modifiable cysteine. The amino acid in the -3 position was the most variable among the lipobox sequences with mainly leucine (71.7%), valine (15.1%), then isoleucine (10.9%) and threonine (2%). In contrast the -2 position was occupied equally by three amino acids serine (38.7 %), alanine (35.4 %), threonine (23.4%) and at low frequencies isoleucine (2.1%), in the -1 position three amino acids residues were found with more constant of glycine (47.3%), alanine (44.2%), and serine (8.3%), and as expected 100% of +1 residue in lipobox was cysteine and the consensus lipobox sequence would be [TIVL][ITSA][SAG][C] as shown in figure 13. The sequence results show that the n-terminal 5-7 residues always contained two high positively charged residues of (lysine or arginine), while the h-region length varied between 7 and 9 residues. Table 12 summarizes the list of 50 lipoproteins *S. aureus* MRSA252 including details of their genes and lipobox structure.

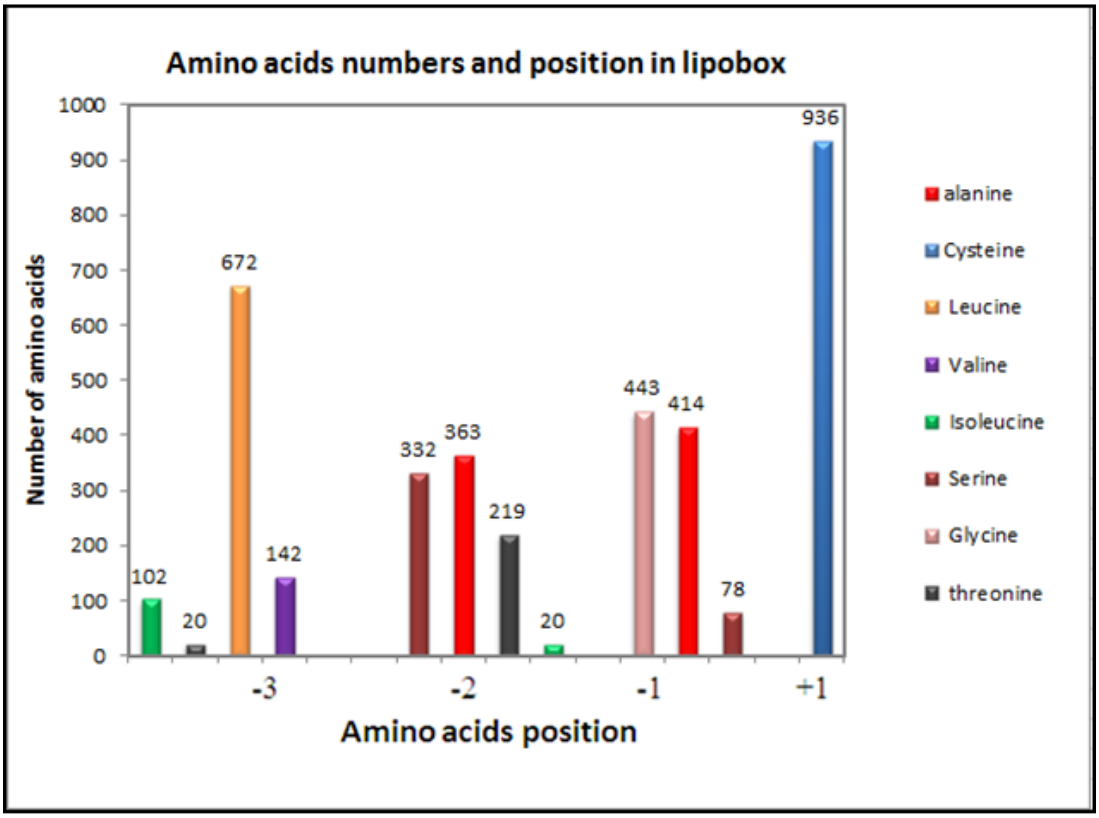

Figure 13. Amino acid numbers and position in lipobox of lipoproteins in 20 *S. aureus* strains

936 lipoproteins were examined for their lipobox properties of 4 amino acid sequences at the C-terminal, the lipid modifiable Cysteine (+1 position) was invariant; Glycine, Alanine and Serine at (-1 position); Alanine, Threonine, Isoleucine and Serine at the (-2 position); Leucine, Valine, Isoleucine and Threonine were occupied at the (-3 position); some minor amino acids were not included.

Table 12. List of *S. aureus* MRSA252 lipoproteins including details of their genes and lipoprotein signal peptides, the carboxy-terminal region (C-region or lipobox), the hydrophobic (H-) region and amino-terminal (N-) region.

ORF	Amino acids sequence length	Gene sequence length DNA (bp)	lipoprotein signal peptides	Lipobox aa length	Lipobox DNA sequences
SAR0118	330	993	n-region: MNKVIK h-region: MLVVTLAFLLV c-region: LAGC	21	TTATTTTGATTGTTTTTCAATATTTAAC TTTTCATATAAATCGTCAATAAGTTTT AATGATGA
SAR0174	324	975	n-region: MKR h-region: LSIIVIIGIFI c-region: ITGC	18	ATGAAAAGGTTAAGCATAATCGTCAT CATTGGAATCTTTATAATTACAGGATG T
SAR0201	591	1776	n-region: MKK h-region: IISIAIIVLALV c-region: LSGC	19	GTGAAGAAAATCATTAGTATCGCAAT TATAGTTTTAGCGTTGGTATTAAGTGG TTGT
SAR0206	423	1272	n-region: MSKILK h-region: YITLAVVMLLI c-region: VTAC	21	ATGTCTAAAATTTTAAAATATATCACG TTAGCCGTGGTAATGTTATTAATCGTA ACTGCATGT
SAR0216	322	969	n-region: MKSK h-region: IYILLLFLIF c-region: LSAC	18	TCATTGAATCATCTCCAAAAATTTATG ATGCGGAATGTCCGGTAATTTAGATTT
SAR0230	491	1476	n-region: MKFKR h-region: LATIFSAVLV c-region: LSGC	19	TTATCGTTCAATCGTAGTTCGATAATC GATTAAATAGATACCTTCAGGTGTTAC TTT
SAR0340	284	855	n-region: MKK h-region: LTTLLLASTLL c-region: IAAC	18	ATGAAAAAGTTAACAACGCTATTATT AGCATCAACGTTATTAATTGCTGCATG T
SAR0390	190	573	n-region: MKLK h-region: SLAVLSMSAVV c-region: LTAC	19	ATGAAATTAAAATCATTAGCAGTGTT ATCAATGTCAGCGGTGGTGCTTACTGC ATGT
SAR0396	208	627	n-region: MKKR h-region: LLLSTFLASTLI c-region:LTGC	20	TTATTTATCGATAACATCACTCTTGAT ACCTTTAGATTTTAAGAAATCTTTAAT TTTATC
SAR0438	261	786	n-region: MMGNIK h-region: SFALYISILLLIVV c-region: VAGC	24	ATGATGGGAAATATAAAAAGTTTTGC ATTGTACATAAGTATCTTGCTTTTAAT AGTTGTTGTAGCAGGTTGT

Table 12-continued

ORF	Amino acids sequence length	Gene sequence length DNA (bp)	lipoprotein signal peptides	Lipobox aa length	Lipobox DNA sequences
SAR0439	257	774	n-region: MGYLKR h-region: FALYISVMILMFA c-region: IAGC	23	ATGGGGTATTTAAAAAGGTTTGCATTG TACATAAGCGTTATGATTTTAATGTTT GCGATAGCAGGTTGT
SAR0442	256	771	n-region: MGYLKR h-region: IGMCISLLIVIIF c-region: VTSC	23	ATGGGATATTTAAAAAGGATTGGAAT GTGCATAAGCCTATTGATTGTAATTAT TTTTGTAACATCTTGC
SAR0443	266	801	n-region: MRYLNR h-region: VVLYIIVMVLSVF c-region: IIGC	23	ATGAGATATTTAAATAGAGTTGTACTG TACATAATTGTTATGGTTTTGAGTGTT TTTATAATAGGTTGT
SAR0444	264	795	n-region: MKSIKR h-region: IGLCISLLILIIF c-region: VTSC	23	ATGAAGTCTATAAAAAGGATTGGATT GTGCATTAGTTTGTTGATTTTAATCAT CTTTGTTACATCTTGT
SAR0463	280	843	n-region: MKR h-region: LIGLVIVALVL c-region: LAAC	18	ATGAAAAGATTGATTGGGTTAGTTATC GTAGCACTTGTATTATTAGCAGCGTGT
SAR0618	295	888	n-region: MKK h-region: SLIAFILIFMLV c-region: LSGC	19	ATGAAAAAAATTGTTATTATCGCTGTT TTAGCGATTTTATTTGTAGTAATAAGT GCTTGT
SAR0641	309	930	n-region: MKK h-region: LVPLLLALLLL c-region: VAAC	18	TTATTTCATGCTTCCGTGTACAGTTTC AATATTTGATTTCATCATTTTGTAGTA
SAR0706	34	105	n-region: MR h-region: FMNEILVHIMTTA c-region: ISGC	19	TTATTTTTTATCACGTTTATGGAGCCA ATAACCAAATAACGTAACGAGACAAC CACT
SAR0730	129	390	n-region: MKK h-region: LIISIMAIMLF c-region: LTGC	18	ATGAAGAAATTAATCATCAGCATTAT GGCGATCATGCTATTTTTAACAGGTTG T
SAR0790	342	1029	n-region: MKK h-region: TVLYLVVAVMFL c-region: LAAC	19	ATGAAGAAAACAGTCTTATATTTAGT AGTAGCAGTAATGTTTTTATTAGCGGC ATGC

Table 12-continued

ORF	Amino acids sequence length	Gene sequence length DNA (bp)	lipoprotein signal peptides	Lipobox aa length	Lipobox DNA sequences
SAR0794	292	879	n-region: MKK h-region: IVIIAVLAILFVV c-region: ISAC	20	ATGAAAAAAATTGTTATTATCGCTGTT TTAGCGATTTTATTTGTAGTAATAAGT GCTTGT
SAR0872	273	822	n-region: MKK h-region: LFGLILVLTFAVV c-region: LAAC	20	ATGAAAAAATTATTTGGTCTTATTTTA GTATTAACATTTGCAGTTGTATTAGCA GCTTGC
SAR0953	551	1656	n-region: MTRKLK h-region: TLILIFVATIA c-region: LSGC	21	ATGACAAGAAAGTTAAAAACGCTGAT TTTAATATTTGTTGCCACAATTGCATT AAGTGGTTGT
SAR1011	319	960	n-region: MNRNIVK h-region: LVVFMLILVVA c-region: VAGC	22	GTGAATAGGAATATCGTTAAATTAGTT GTGTTTATGCTAATCTTAGTTGTAGCA GTAGCGGGTTGT
SAR1034	366	1101	n-region: MSKFK h-region: SLLLLFGTLIL c-region: LSGC	20	TTAATGTCCACCTCCATGATCATCATT GTCTTGATCTTGCGCATCTTTTGAAAT TTTCTT
SAR1066	208	627	n-region: MKFGK h-region:TIAVVLASSVL c-region: LAGC	20	ATGAAATTTGGAAAAACAATCGCAGT AGTATTAGCATCTAGTGTCTTGCTTGC AGGATGT
SAR1106	292	879	n-region: MRIIK h-region: YLTILVISVVI c-region: LTSC	20	TTGAGAATCATAAAGTATTTAACCATT TTAGTGATAAGCGTCGTTATCTTAACC AGCTGT
SAR1189	317	954	n-region: MKK h-region: TLGCLLLIMLLV c-region: VAGC	19	ATGAAAAAGACACTGGGATGTTTACT TTTAATTATGCTTTTAGTCGTAGCAGG TTGT
SAR1288	118	357	n-region: MRR h-region: WFVLILGLVIL c-region: LSAC	18	TTATTTTTTCATATTATTTTCTTCATAA ACTGGTTTATTATCTTTCTCAAACTT
SAR1494	302	909	n-region: MFKRTK h-region: LILIATLL c-region: LSGC	18	TTAGTCAAACACACCATATTCTGAAGT GCTTTTACCGTTATAATTTGGTTTAAC

Table 12-continued

ORF	Amino acids sequence length	Gene sequence length DNA (bp)	lipoprotein signal peptide	Lipobox aa length	Lipobox DNA sequences
SAR1495	301	906	n-region: MLKKAK h-region: LILIATLL c-region: LSGC	18	TCATTCGACCTCAATCCTTATAGACTC ATTATCACTGTAATTAACTCGATTAGT
SAR1558	144	435	n-region: MKK h-region: VIGLLLVSTLA c-region: LTAC	18	ATGAAAAAAGTAATCGGACTGCTACT AGTAAGTACATTAGCTTTAACAGCTTG T
SAR1608	193	582	n-region: MKK h-region: LVSIVGATLL c-region: LAGC	17	ATGAAAAAATTGGTTTCAATTGTTGGC GCAACATTATTGTTAGCTGGATGT
SAR1831	281	846	n-region: MKK h-region: LIFLIVIALV c-region: LSAC	17	TTGAAAAAGTTAATATTTTTAATTGTA ATTGCTTTAGTTTTAAGTGCATGT
SAR1879	184	555	n-region: MLK h-region: GCGGCLISFIILIIL c-region: LSAC	22	ATGTTAAAAGGATGCGGCGGTTGCCTT ATTTCTTTTATTATATTAATTATCTTAT TATCAGCCTGT
SAR1881	208	627	n-region: MKFK h-region: AIVAITLSLSL c-region: LTAC	19	ATGAAATTCAAAGCTATCGTTGCAATC ACATTATCATTGTCACTATTAACCGCC TGT
SAR1995	399	1200	n-region: MKR h-region: TLVLLITAIFI c-region: LAAC	18	TTAATTACTGTAAATATGAACTTGCGG TTCTTTGTCATCTTTTGTCTTACTAAT
SAR2104	154	465	n-region: MKR h-region: LLGLLLVSTLV c-region: LSAC	18	ATGAAAAGATTGTTAGGTTTATTATTA GTGAGCACGTTAGTGTTAAGTGCATGT
SAR2179	290	873	n-region:MKKK h-region: TLLPLFLGIMVF c-region: LAGC	20	TTATTTCTTTTTCTTTTTAGACACTACT TGTGTGTTTTTGCCTTTTTTATTGCTGC CGCC
SAR2268	327	984	n-region: MRGLK h-region: TFSILGLIVALFL c-region: VAAC	22	TTACTTTTGTTCTTTTTTTGATAATTCA ACAAGTTCTTTAGCCATTTCTTCAGAA GAAATTAAGCC

Table 12-continued

ORF	Amino acids sequence length	Gene sequence length DNA (bp)	lipoprotein signal peptides	Lipobox aa length	Lipobox DNA sequences
SAR2363	260	783	n-region: MKMKR h-region: FIAIVMALFLV c-region: LAGC	20	TTATGCTGTAAAGTGGTATTCTTTTAA TATTTCTTTAGCTTTATCTGATTTTAAG AATTC
SAR2368	302	909	n-region: MKK h-region: LLLPLIIMLLV c-region: LAAC	18	TTATTTTGCAGCTTTAATTAATTTTTCT TTTAAATCTTTACGCATGAAATCTAA
SAR2457	209	630	n-region: MKK h-region: LVTGLLALSLF c-region: LAAC	18	TTATTGTTGGTAGTTTGGATCAGTAACC ATTGCTTGTCCAGTATAATCAACCGT
SAR2496	515	1548	n-region: MKKK h-region: LGMLLLVPAVTLS c-region: LAAC	21	TTAATGCGCTAACATTTCTTCTTTGATT TGGTCTTTATTTAATTTTGAAGGATAAT ATGTTGG
SAR2499	199	600	n-region: MTKK h-region: LLTLFIVSMLI c-region: LTAC	19	CTATTTGATTTTATCTTTTAATAATTTTT CATAACTTTCATAATCATATGGATCTTC
SAR2500	120	363	n-region: MKR h-region: FVATVLLLLVF c-region: ISGC	18	CTAGTTCGTCATATTTTCTTCATGATAAT CCGCATCATGTTGTTGATAATATCG
SAR2504	259	780	n-region: MKR h-region: LLFVVIAFVFI c-region: LAAC	18	CTATTTAGATTTAGAAACATCTTGACCA AACCATTTCTTACCTATTTTAGCTAA
SAR2536	313	942	n-region: MKKIK h-region: YILVVLVLSLTV c-region: LSGC	21	TTACTTATGACCACCTTTCTGTTTATCA AAATAGTGGTGTTTTTTCAAAAATTCT TCTGCGAC
SAR2546	151	456	n-region: MKK h-region LCSLIVVALVCVIA c-region: LSAC	21	TTGAAAAAATTATGTTCATTAATTGTAG TAGCATTAGTTTGTGTCATTGCATTATC AGCTTGT
SAR2554	532	1599	n-region: MRKLTK h-region: MSAMLLASGLI c-region: LTGC	21	TTATTTATACTGCATTTCATTGAATGGTA ATTCATACTGTGATTGTGTGAATGCTAC TTTTTC

3.2.5 Phylogenetic analysis

The phylogenetic trees generated with both analytical programs package (Seaview 4 and MEGA 6) using the lipoprotein sequence data revealed very similar results to confirm the accuracy of these analyses. Concatenated sequences were used to construct a Bayesian phylogenetic tree of the 44 common lipoprotein genes (figure 14 A and B). Six genes were identified as not occurring in most strains or with truncated sizes (*SAR1558, SAR1288, SAR0439, SAR0706, SAR1831* and *SAR2104*) were excluded in the phylogenetic analysis. Phylogenetic analysis was completed by using a Neighbor-Joining (NJ) method, based on pairwise nucleotide sequence alignments for the 44 *S. aureus* lipoproteins. Both trees in figure 14 had similar distributions of 20 *S. aureus* strains according to their lipoproteins, based on the genetic diverse and lipoproteins variations. The clusters of these strains were highly concordant with BLAST alignment data of individual genes. There were two main branches for both trees dividing to two separate groups of strains (8 and 12 strains for each group) figure 14 A and B, these strains showed clear relation between their lipoprotein genetic structure, each group had two or three sub-groups of more closely related strains. The first group of lipoprotein in both trees has sub-groups were almost distributed in pairs to show more close related strains, three groups of pair strain are generated (MRSA252 and T1), (MSSA476 and MW2) and (RF122 and M013) these strains shared almost identical genetic features for lipoprotein genes.

RF122 bovine strain and M013 MRSA strain showed more than 98% average identity of examined genes nucleotide sequences, but *SAR0438* shared only 88% of DNA identity, also 96% of the *SAR0444* and *SAR0730* sequence was identical in both strains. MSSA476 and MW2 strains displayed homologous lipoprotein genes, this high similarity sequence were found in all genes. The related MRSA252 and T1 strains have shown a strong genetic structure similarity for the tested genes to suggest the common genomic features, meanwhile, LGA251 and ST398 strains are possess less nucleotides identity with MRSA252 to show earlier diversion according to the percentage of similarity as shown in figure 14.

On the other hand, the trees in figure 14 A and B has second branch of were distributed into on other two main divisions one of them were contains three HA-MRSA strains (Mu3, Mu50 and N315) with highly related lipoprotein primery sequence, the three strains were sharing the similar nucleotide identity with ~100% identity, there were three lipoprotein

genes (*SAR0439*, *SAR0438* and *SAR1011*) have very minor nucleotide differences in Mu3 strain, for this reason there is a slight variations in this cluster. The second sub-branch consisting of 9 strains distributed in individual and pairs, HA-MRSA strains T0131, BMB9393 and TW20 has similar nucleotides properties close to 98%, however T0131 and BMB9393 were more closely related to each other, by cause of few variable genes (*SAR0439*, *SAR0444*, *SAR0790*, *SAR042* and *SAR0443*) that contains more nucleotides variation to generate this differences. The laboratory strain NCTC8325 and vancomycin resistant VC40 strain were shown conserved relative nucleotides structure, both strains were identified with 100% identitical in 44 lipoprotein genes, while the MRSA strain COL had two diverse genes (*SAR0439* and *SAR0442*) with 2% and 8% nucleotides variability amongb these genes, this nucleotides variability produce a slight alteration in this small cluster group. Additionally, MRSA *S. aureus* M1 have almost the similar nucleotides features of this cluster but had a few further genetic differences in *SAR0438* and *SAR0442* genes with some variant SNPs. Unlike most other strains, *S. aureus* str. Newman and RN4282 had a different trend were branched in an individual manner to show less nucleotides identity to the other groups.

(A) **(B)**

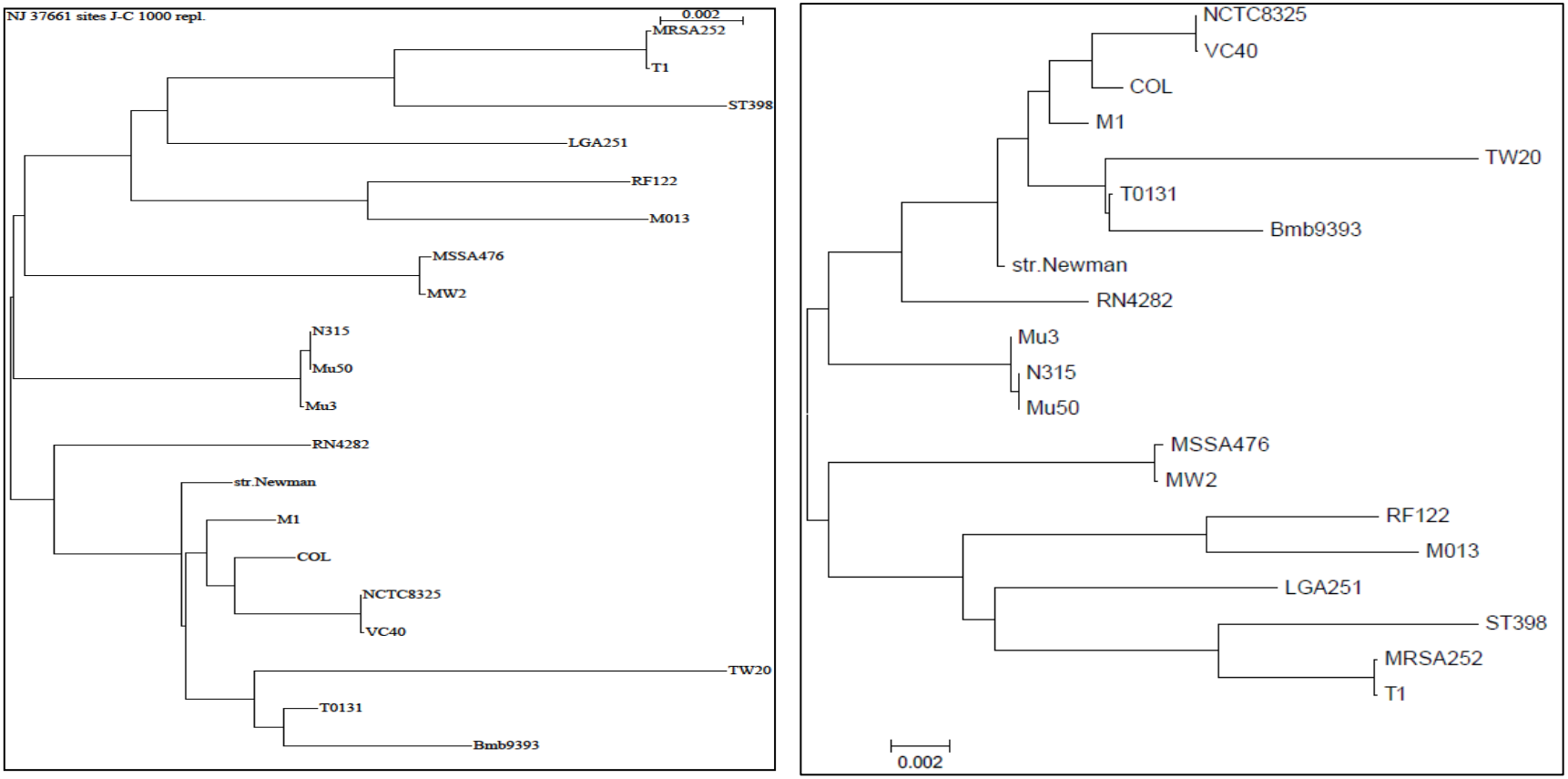

Figure 14. Phylogenetic relationships of lipoprotein among 20 *S. aureus* strain

Bayesian phylogram indicating the evolutionary relationships of lipoprotein, sequences were aligned using Clustal W2. NJ tree with bootstrapping consensus inferred from 1,000 replicates was constructed using sequences of concatenated 44 lipoprotein genes for each *S. aureus* strain representing 37670 bp by two different software (A) MEGA 6, and (B) Seaview 4 were used to generate the concatenated trees.

3.2.6 Antibiotic susceptibility test

In this study *S. aureus* strains has been used in the proteomics and genomics analysis were investigated to determinate their methicillin resistance using oxacillin/cefoxitin disc diffusion test on Iso-Sensitest agar. The variability of MICs of *S. aureus* strains could help to explain the differences in some genetics structures and expression levels during the growth stages. Results are presented in table 13. Three strains (MRSA252, RN4282 and 8325-4) were resistant to cefoxitin according to current criteria and were therefore probably MRSA strains, while only 1 strain (T1) was susceptible oxacillin/cefoxitin antibiotics. Also the full antibiotic susceptibility test images for each strain prepared in agar plates are available in appendix B.

Table 13. Zone diameter and MIC interpretive for *S. aureus* tested with 30 µg of cefoxitin

MIC breakpoints to determined inhibition zone: susceptible = ≥ 22 mm diameter, resistant = ≤ 21 mm diameter

S. aureus strain	Zone diameter interpretive criteria (nearest whole mm)
MRSA252	0 mm
T1	25mm
RN4282	21mm
ACTC8325	21mm
NCTC12493 (control strain)	10mm

3.3 Discussion

3.3.1 Lipoprotein relatedness of *S. aureus* strains and their genetic diversity

S. aureus strains have a genome ranging in size from 2,799,802 to 3,075,806 base pairs containing regions of small and large difference. Small genetic variation is defined as changes that affect few individual genes, while the large variation is related to horizontally acquired DNA segments. Different predictive algorithms rules have been used to predict the unique N-terminal lipid modification. The absence of some genes gives a clear indication that these genes are not common to all *S. aureus* strains, furthermore the functions for most of them were not well identified to be named as putative lipoproteins. Also the comparative analysis search for lipoproteins in protein database showed the numbers of bacterial lipoproteins were not similar between strains (Sutcliffe and Harrington, 2002). The BLAST search results in this study have revealed certain variation in the lipoprotein repertoire among examined strains. Six genes were absent in some tested strains. Lipoprotein genes of T1 strain which has revealed to be closely related to other lipoprotein genes of MRSA252, but *SAR1881* was not found in T1 *S. aureus*. Also, small number of genes (*SAR2104, SAR1881, SAR1494* and *SAR0953*) in RN4282 and T1 strains were absent from MRSA252.

Gene sequence analysis of nucleotide polymorphism was used to detect SNPs between lipoprotein of *S. aureus*, revealed 11 genes with strong nucleotides changes leading to many amino acids variations, these variant genes have a strong impact on the phylogenetic trees as some strains was more related to each other based on the sequence variation of these genes. From molecular analysis, the great majority of mutations among lipoprotein genes are represented within a limited number of genes to reveal considerable variation among analogous genes encoding lipoproteins, 11 highly variable genes with considerable numbers of non-synonymous residues.

3.3.2 Phylogenetic analysis

The most reliable markers genes used in epidemiological or population studies encode essential housekeeping functions (Case *et al.*, 2007). A study in population and epidemiological investigation based on ~17.8 kb of intra-species tree sequence of individual genes using multilocus sequence typing (MLST) suggested that the candidate marker loci used in *S. aureus* phylogeny should have at least >1 % mean pairwise diversity

for all genes in the genome, also there was no strong relation between gene function and phylogenetic reliability (Cooper and Feil, 2006). The neighbour joining tree of *S. aureus* strains investigated in this study indicates a pattern of genetic diversity of lipoprotein, the distribution of this diversity was generated and displayed in phylogenetic trees, *S. aureus* lipoprotein differed from one another among 20 strains, also it has revealed several common features in the majority of examined genes. To illustrate the value of variety in compared lipoprotein sequences with related genes sequences, both analyses trees formed distinct clusters of strains with similarity of 100%. The main two branches of phylogenetic tree has a consistent structure among their sub-branches, 8 strains were trend to share more genetic features and exposed to less mutation changes. Comparisons between UK hospital-acquired epidemic MRSA252 and T1 strains by using BLAST search for the lipoprotein genes, the two strains had an identical genetic identity, lipoprotein genes identity were found 100% identical in each strain, MRSA252 has proved a high methicillin resistance while T1 were methicillin susceptible. LGA251 and ST398 strains have shown more genetics identity related to MRSA252 strain with few minor varability in some genes. Lipoprotein genes sequence of MSSA476 and MW2 strains were highly identical, these results are consistent with previous study which find that both strains was related to each other (Holden *et al*., 2004). Based on the generated phylogeny comparative analysis in this study the genetic structure relationship between MSSA476 and MW2 strains the genetic elements of lipoprotein were highly identical. The sequence and order of most of the open reading frames in the genomes M013 and MW2 were conserved (Huang *et al.*, 2012). Lipoprotein genes of bovine strains RF122 was similar to genes content to human strain M013, this finding suggests that similar strains may cause cross infection between human and animals.

NCTC 8325 and VC40 strains have shown major paired nucleotides structures in all tested genes, this high similarity were proved in other genetic study were genomic tree based on average nucleotide identity (ANI) calculation of homologous genes among five *S. aureus* genomes has shown that VC40 strain were closely related to strains NCTC 8325 (Kim *et al.*, 2014). A previous study using an amplicon-based microarray technique found that COL and 8325 stains were clustered closely to show very identical related genomically features (Cassat *et al.*, 2005). MRSA N315, Mu3 and VRSA Mu50 strains genome content share common characteristics, complete comparative analyses of whole genome sequences of three strains found 213 ORFs diverse between Mu50 and N315, and 9 ORFs between

Mu50 and Mu3 (Ohta *et al.*, 2004). N315 and Mu50 strains share 96% nucleotide sequence identity (Kuroda *et al.*, 2001). TW20 strain that is highly resistant to a range of antibiotics shares close average nucleotide identity with T0131and BMB9393 strains. Comparative study on the genomes of these *S. aureus* shown that BMB9393 has 142 unique ORFs, also BMB9393 and TW20 strains sharing 2,555 CDS and 2,541 with strain T0131 (Costa *et al.*, 2013). T0131 shares 2,573 orthologous coding sequences CDSs with TW20, moreover T0131 has 85 definite genes compared with TW20 genomes (Li *et al.*, 2011).

Baba *et al*., 2008 examined the diversity and clonal relationships between the *S. aureus* strains to investigate the differential success of these strains based on sequence of seven housekeeping genes and the distribution of the genomic islands (Baba *et al*., 2008), data has revealed phylogenetic analyses identical to the overall results of phylogenetic trees in this study, the shared tested strains in both studies were grouped in the same pattern and groups. The sequence variation of lipoproteins within virulent strains may have a value as an epidemiological marker among related isolates, but the results for most lipoprotein genes in this study had <1 % mean pairwise diversity to indicate their low significance as genetic markers.

The lipoprotein genes *SAR0439* putative lipoprotein and *SAR0442* putative membrane protein have shown the highest nucleotide diversity value but it is still insufficient to use as epidemiological marker, as the marker loci used in *S. aureus* phylogeny should have at least >1 % mean pairwise diversity for all genes in the genome.

3.3.3 Lipobox features

The lipobox was identified as a consensus sequence in prolipoproteins, this signature can help to differentiate lipoproteins and non-lipoproteins (Babu and Sankaran, 2002). The lipoprotein signal sequences are characterized with an n-region including 5 to 7 amino acids with at least one or two positively charged amino acid Lys and Arg residues, C-terminal signal recognition sequence containing conserved lipid-modified Cys at +1 position, while the length of uncharged h-region ranges between 7-22 amino acids. Lipobox motif composition of examined 20 *S. aureus* strains did not display major differences and also showed only a small number of varations (29 mutations in 936 distinct lipoboxes) among their amino acids residues. Comparison of the main lipobox features for tested *S. aureus* strains comprised [TIVL][ITSA][SAG][C], in contrast lipobox from other

bacteria [ILV][ASTVI][GAS][C] (Babu and Sankaran, 2002), this residues specifying few amino acids sequences variability, the results showing a slight changes in the amino acids frequency order, lipobox of *S. aureus* have some diverse amino acids residues with a low frequency that were present in some lipoproteins. SNPs occurred in a few lipoprotein genes but only twenty nine amino acids mutations were reported in 8 lipoproteins to show less than 0.85% of all examined lipoboxes. Also a few amino acid residues were minor components in the lipobox as this was noted in *SAR1558* the -3 position was threonine, while in *SAR0443* the -2 position was isoleucine.

3.3.4 Antibiotic susceptibility test

Cefoxitin is recommended to detect the methicillin resistance in *S. aureus* (MRSA), the aim was to evaluate the *S. aureus* strains methicillin resistance for the five strains that been used in this study to assess any possible relation of methicillin resistance and lipoprotein features of the tested strains. Relating to cell wall biosynthesis, multiple genes appear to be significantly induced by different cell wall-active drugs or antibiotics. In general, beta-lactam antibiotics targeting the transpeptidase of PBPs, this action decreases the cell wall cross-linking and hence the lack of cell wall integrity (Mallorquí-Fernández *et al.*, 2004). Lipoprotein genes of MRSA252 and T1 strains were highly similar in the primary sequence, however MRSA252 has proved a highest cefoxitin resistance in the same time T1 strain was highly susceptible. T1 strain was isolated from a case of menstrual TSS in the early 1990's, at that time the methicillin-resistant *S. aureus* strains had not spread so widely. The data shows that the lipoproteins of the two strains are similar suggesting that MRSA252 may have originated from the same lineage as T1. Mobile accessory genetic elements containing virulence genes possibly move between isolates, these acquired elements could explain the relatedness between *S. aureus* strains to shown relative variable MIC breakpoints of inhibition zone for different antibiotics (Moore and Lindsay, 2002). Many transcriptomic experiments on *S. aureus* have defined the existence of different genes whose expression is changed after exposure to a range of antimicrobial agents that target cell wall biosynthesis (Dengler *et al.*, 2011, Drummelsmith *et al.*, 2007, McAleese *et al.*, 2006). Genetic and biochemical examinations carried out on 30-kDa lipoprotein prsA of methicillin resistant *S. aureus* showed that a *prsA* deletion led to changes in the susceptibility of *S. aureus* COL (MRSA) to show decreases the resistance to oxacillin (Jousselin *et al.*, 2012).

Chapter four

Proteomics identification of *S. aureus* lipoproteins under different growth conditions

4. Proteomics identification of *S. aureus* lipoproteins under different growth conditions

4.1 Introduction

Understanding of the lipoprotein properties in *S. aureus* is needed for better definition of their pathogenesis and development of potential therapeutic strategies to reduce *S. aureus* infections. Lipoproteins have been described that are very important virulence factors as they are involved in many processes of *S. aureus* pathogenesis, but fulfil diverse functions. The roles of individual lipoproteins have not been fully characterized. Although some lipoproteins are involved in Fe transport and are regulated by Fe via fur, the regulation of expression of the majority is not understood. Initial studies therefore aimed to determine whether the lipoprotein genes were expressed in different growth phases. Investigation of lipoproteins could help to understand mechanisms of virulence and antibiotic resistance of *S. aureus*. Many transcriptional analyses of proteins expressed on *S. aureus* have provided views on the virulence factors expression under different conditions, but these results cannot explain the proteome changes during host infection as production of various proteins can be regulated post-translationally. Just because the gene is present and transcribed does not prove that the protein is synthesised. A detailed proteomic profiling of *S. aureus* lipoproteins is important to confirm their expression from the genes and to determine their potential roles as virulence factors and antigens. This proteomic information could help to identify the differences and similarity of lipoprotein constituents and provide helpful information to develop a vaccine and/or treatment of *S. aureus* infections.

There are two major approaches to convert proteins extracted from a biological specimen to peptides prior to mass spectrometry and proteome analysis. The first approach is to treat proteins with detergents, separate them by sodium dodecyl sulfate (SDS) poly acrylamide gel electrophoresis (PAGE) followed by in-gel digestion of proteins (Gorg *et al.*, 2004). The second approach is treating the protein mixture with strong chaotropic substances e.g. urea. Protein digestion under detergent-free conditions is known as in-solution digestion gel-free proteomic approaches can be also divided as bottom-up and top-down, the bottom-up is to identify corresponding proteins by analysis of peptides obtained via specific proteolysis. While, the top-down proteomics is based on using MALDI-based tandem mass spectrometry of fragments or intact protein for identification (Yates *et al.*, 2009).

Hecker *et al*. summarized the impact of proteomic methodology and mass spectrometry approaches to understand the *S. aureus* mechanisms of pathogenesis and virulence (Hecker *et al.*, 2010). The classical methods mass spectrometry (MS) or tandem mass spectrometry (MS/MS) for quantitative analysis of primary sequence and posttranslational modifications have shown successful results when applied to small amount of proteins (Mann and Jensen, 2003). Mass spectrometry is a powerful technique for high-throughput proteome analysis based on the mass of peptides subjected to trypsin digestion (Aebersold and Mann, 2003). The qualitative two-dimensional gel electrophoresis 2-DE in combination with protein and peptide analysis mass spectrometry methods can identify individual protein spots. This method is based on protease digestion of protein mixtures and subsequent peptides separated and identified by LC-MS/MS (Lim *et al.*, 2003).

The development of non-gel based proteomic techniques has provided powerful tools to investigate protein quantification on a large-scale and characterisation of proteins in complex biological models (Cravatt *et al.*, 2007). Label-free quantification methods correlate the MS signal of peptides obtained by trypsin digestion of proteins or number of peptides sequenced with the relative or absolute protein quantity. These quantitative approaches can measure significant proteomic changes within a complex proteins mixture of multiple samples within a single experiment with a high analytical depth and dynamic range. Each sample is prepared separately then individual samples subjected to LC-MS/MS runs and data analysis including peptide/ protein identification, quantification and statistical analysis (Old *et al.*, 2005). The label-free shotgun proteomics techniques for quantitative proteomics methods provide higher dynamic range of proteins quantification. There are two different label-free quantification strategies have been used, the first one is measuring and comparing MS signal intensity of particular peptide precursor ions belonging to a specific protein, while the second strategy of label-free LC-MS quantification measurements to perform an individual LC-MS/MS runs and protein abundance changes are calculated by counting and comparing the number of each spectra between different experiments (Chelius and Bondarenko, 2002).

The aim of this section of the study was initially to identify and characterise all expressed lipoproteins of *S. aureus* MRSA252 an epidemic MRSA strain causing hospital-acquired infections in United Kingdom. In this chapter, the first step was proteomic profiling analyses of *S. aureus* lipoproteins by determining and identifying the lipoproteins via matrix-assisted laser desorption ionization-time of flight MALDI/TOF and liquid

chromatography-mass spectrometry LC-MS/MS analysis methods. Furthermore, an analytical quantitative experiment was performed to present a comprehensive quantitative proteome profiling of *S. aureus* MRSA252 lipoproteins, complementing the method of gel-free/in-solution trypsin digestion proteins.

4.2 Results

4.2.1 1D and 2D protein separation (procedure 1)

In order to identify and characterise the lipoprotein complement of *S. aureus* MRSA252, bacteria were grown in tryptone soya broth under non-infection conditions and proteins extracted were in time different times of bacterial growth. Proteins separation was performed by two approaches, 1D and 2D gel electrophoresis. The first step was to separate complex protein mixtures extracted from cells treated with 2% Triton X-114 detergent followed by fractionation by 1D SDS-PAGE gel electrophoresis using 12% acrylamide gel as described in section 2.3.2.1. The gel was sliced into sections containing individual bands. Proteins in gel were then reduced, alkylated and digested using trypsin and peptides identified by LC-MS/MS. By the procedure of using 1D SDS-PAGE in combination with in-gel trypsin digestion of proteins in excised gel pieces revealed that few of the identified proteins were lipoproteins, peptide masses were searched by applying Mascot software using a protein database comprising of *S. aureus* MRSA252 protein sequences extracted from the UniProt KB/Swiss-Prot database, 1D gel-based proteomics detected only 6 lipoproteins of the most highly abundant proteins based on the number of identified peptides, lipoproteins identified are presented in table 14 together with other identified proteins detected in this experiment which were not lipoproteins. From the Mascot search results the detected protein scores were presented with the sequence coverage method which correlate the identified peptides to proteins by using the MOWSE (Molecular Weight Search) databases, proteins which fell under the p-value level of ($p<0.05$) were not considered as statistically significant. Overall there were 6 lipoproteins with a high significant p-value and matches to limit individual score (table 14).

Some modifications to the isolation procedure were made to the isolation protocol to try to improve the recovery of lipoproteins including the addition of the non-ionic Triton X-114 detergent before the phase separation, (initially the Triton X-114 was added to the bacterial mixture before the centrifugation at 4°C for 10 min) and discarding the insoluble materials and unbreakable bacterial cells, the modified method used was designed to spin and discard the insoluble materials before the Triton X-114 mixing.

Two-dimensional gel electrophoresis (2-DE) was performed using isoelectric focusing (IEF) with immobilized pH gradients (IPGs) in the first dimension and sodium dodecyl sulfate (SDS-PAGE) for the second dimension and was used to separate lipoprotein

preparations. Individual protein spots in gels were subjected to in-gel trypsin digestion followed by analyses on peptide mass fingerprinting MALDI/ TOF- MS. The results from 2-DE approach once again showed insufficient separation of lipoproteins and only three lipoproteins were detected in the majority of gels as presented in table 15. These were different to the lipoproteins detected by the previous method. The probability that the observed match is a random event and protein scores greater than 82 are significant ($p<0.05$), Mascot threshold was set at $P<0.05$, threshold Mascot score given corresponds to probability of 95% of the identified protein is not a random match.

Six of these lipoproteins showed a high abundance during all three times of label-free relative protein quantification experiment. On the other hand, the 2-DE gel approach showed a similar poor efficiency of both quantity and quality of lipoprotein detection. In this procedure the lipoprotein spots in the second dimension gels were not stable in the same distribution arrangement in the different sizes of immobilized pH gradients (IPGs) strips, both sizes pH 3.5-10 NL 18 cm and pH 3,5 -6 NL 7 cm IPGs strips had a different pattern of proteins separation which made it hard to differentiate between the identical spots in different gels sizes.

Only three lipoproteins (SAR0641, SAR2457 and SAR2504) were identified in high abundance with the 2-DE gel approach, these proteins were detected in individual spots in different gel sizes.

Table 14. List of identified lipoprotein of *S. aureus* MRSA252 in 1D SDS-PAGE 1D

SDS-PAGE combined with in-gel trypsin digestion followed proteomic analysis by LC-MS/MS and Mascot search.

Gene ORF	UniProtKB accession No	Protein description	MW (kDa)	Total score	Sequence coverage %	Sequence matches
SAR0390	Q6GJS7	Hypothetical lipoprotein	21.35	328	37	8
SAR0641	Q6GJ39	ABC transporter	34.77	421	29	23
SAR2504	Q6GE19	Extracellular solute-binding lipoprotein	28.85	317	43	14
SAR0872	Q6GIH7	Lipoprotein	30.36	302	25	13
SAR2368	Q6GEF0	Ferrichrome-binding lipoprotein	34.04	290	29	14
SAR0216	Q6GK91	Putative lipoprotein	38.21	187	24	8
Proteins detected in this experiment were not lipoprotein						
SAR1157	Q6GHQ4	Penicillin-binding protein	82.72	216	29	11
SAR0566	Q6GJA7	Putative surface-anchored protein	98.62	311	21	19
SAR2580	Q6GDU5	Fibronectin-binding protein precursor	105.7	201	31	7
SAR0553	Q6GJC0	Translation elongation factor Tu	43.11	189	33	10
SAR2544	Q6GDX8	ABC transporter amino acid-binding protein	25.16	299	26	15
SAR1232	Q6GHH9	Ribosomal protein S2	29.1	170	24	11

Table 15. List of identified lipoprotein of *S. aureus* MRSA252 in 2-DE

2-DE combined with in-gel trypsin digestion followed by proteomic analysis via LC-MS/MS and Mascot search.

Gene ORF	UniProtKB accession	MW (kDa)	Total score	Calculated pI	Sequence coverage (%)
SAR0641	Q6GJ39	34.75	267	8.3	83%
SAR2457	Q6GE63	23.33	299	6.06	89%
SAR2504	Q6GE19	28.82	251	9.2	79%

4.2.2 Protein and peptide quantitation using LC-MS/MS methods

To investigate the expression of lipoproteins in various bacterial growth stages a comprehensive proteomic analysis of *S. aureus* MRSA252 lipoprotein was performed. Bacteria were grown in tryptone soya broth and incubated at 37°C with constant shaking at 150 rpm. Cultured bacterial cells were harvested after 15 h (late exponential growth phase) then subjected to detergent extraction by using the non-ionic Triton X-114 phase separation to solubilize lipidated proteins, followed by gel-free in-solution protein digestion with trypsin and analysis of generated peptides via combining approach of C18 ion pair of peptide chromatographic separation with MS/MS analyses, the data is based on analysis of two biological replicates. In this experiment, mass spectrometric analysis using the Mascot software and UniProtKB database revealed the identification of a total of 38 distinct lipoproteins as listed in table 16. Twenty three of these proteins were identified as putative lipoproteins. Thirty two lipoprotein detected were in the 50 DOLOP database lipoprotein list of *S. aureus* MRSA252 and 6 lipoproteins were not. Protein scores were different among the detected lipoproteins with the highest score of 4122 was a 21.3 kDa putative lipoprotein SAR0390, while the lower-protein-score of 83 was a 29.7 kDa putative lipoprotein SAR0106.

Table 16. List of *S. aureus* MRSA252 lipoproteins identified in gel-free in-solution proteins digestion method and analysed by LC-MS/MS

Peptide identification was accepted if established at >95% probability.

Gene ORF	Protein description	MW (kDa)	Protein score	Protein matches significance	Protein sequences significance
SAR2457	Putative lipoprotein	23.36	1052	28	8
SAR0445	Putative lipoprotein	31.66	431	11	9
SAR0443	Putative lipoprotein	30.90	201	5	4
SAR0438	Putative lipoprotein	30.32	237	6	5
SAR0340	Efem/EfeO family lipoprotein	32.28	158	4	5
SAR0106	Putative lipoprotein	29.78	83	3	3
SAR2763	Putative lipoprotein	43.37	517	18	11
SAR2546	Putative lipoprotein	17.30	641	15	11
SAR2504	Extracellular solute-binding lipoprotein	28.85	3652	87	25
SAR2500	Putative lipoprotein	14.37	453	16	8
SAR2499	Putative lipoprotein	23.11	168	7	4
SAR2496	Putative solute-binding lipoprotein	59.16	217	5	3
SAR2368	Putative ferrichrome-binding lipoprotein	33.98	1714	44	20
SAR2268	Putative transport binding lipoprotein	36.64	492	20	15
SAR1995	Putative lipoprotein	45.43	789	21	16
SAR1881	Putative lipoprotein	23.80	240	6	4
SAR1608	Putative lipoprotein	21.51	414	9	7
SAR1558	Putative lipoprotein	16.04	257	4	2
SAR1494	Putative lipoprotein	34.48	352	7	5
SAR1288	Putative lipoprotein	14.17	1406	45	11
SAR1189	Putative lipoprotein	35.99	507	13	10
SAR1066	Putative lipoprotein	23.90	332	7	6
SAR0953	Transport extracellular binding lipoprotein	61.52	119	7	5
SAR0872	Lipoprotein	30.36	2141	79	22
SAR0839	Putative lipoprotein	28.44	355	16	13
SAR0794	Putative lipoprotein	34.09	435	14	10
SAR0761	Putative lipoprotein	16.14	560	20	11
SAR0730	Putative lipoprotein	14.99	627	42	21
SAR0641	ABC transporter extracellular binding protein	34.77	3695	138	31
SAR0618	Putative transport system lipoprotein	33.27	109	2	2
SAR0463	Lipoprotein	30.49	817	33	17
SAR0396	Putative lipoprotein	23.70	414	10	7
SAR0390	Putative lipoprotein	21.35	4122	121	28

Table 16-continued

Gene ORF	Protein description	MW (kDa)	Protein score	Protein matches significance	Protein sequences significance
SAR0230	Extracellular solute-binding lipoprotein	55.51	114	9	8
SAR0216	Putative lipoprotein	36.89	425	29	16
SAR0206	Extracellular sugar-binding lipoprotein	47.92	148	8	6
SAR0174	Putative lipoprotein	36.54	341	18	12
SAR0145	Putative lipoprotein	35.08	682	14	9

4.2.3 Differential lipoprotein expression changes (procedure 2)

As the gel-free in-solution digestion followed by LC-MS/MS was able to resolve the lipoproteins a second set of experiments was performed to quantify and differentiate lipoproteins of *S. aureus* grown in 3 different culture conditions using both quantitative and qualitative techniques to compare the different samples. The relative quantification method based on simple data analysis of expression fold changes ratio was used. This gel-free proteomics provides an overview on the comparative quantitative lipoprotein of *S. aureus* MRSA252. The abundance of peptides in the label-free approach was measured based on the identification of peptides by LC-MS/MS and used the total number of tandem mass spectra that matched peptides to a particular protein to quantify the abundance of lipoproteins in the samples. Samples were then statistically compared to evaluate the differences between the three conditions. In order to illustrate the possible expression similarities between the lipoprotein of *S. aureus* a comparative analysis was performed for three different bacterial growth phases early exponential (6h), late exponential (15 h) and stationary phase (24 h) as shown in figure 15, the data was based on analyses of two biological replicates. Identification and analyses of peptides quantitation revealed a quantitative expression level for 38 lipoproteins of *S. aureus* MRSA252. Detection and comparison of 38 lipoproteins corresponds to coverage of approx. 70% of predicted *S. aureus* MRSA252 lipoproteins. The detected lipoproteins and their regulation expression ratios are summarized in table 17. There were some differences in lipoprotein expression between the late exponential and stationary phases, 17 lipoproteins showed lower production in the stationary phase, the expression comparison between the stationary and

late exponential phases showed that SAR0443, SAR0106, SAR2546, SAR2504, SAR2499, SAR2496, SAR1608, SAR1189, SAR1066, SAR0953, SAR0839, SAR0794, SAR0618, SAR0396, SAR0216, SAR0206 and SAR0174 were down-regulated after 15 h growth. SAR0206 and SAR2763 lipoproteins were not produced at 6 h but at the late exponential phase had a moderate expression level. Lipoproteins synthesized in the stationary phase showed down-regulation for 14 lipoproteins, SAR2763 lipoprotein was not expressed in the stationary phase, however, 20 lipoproteins displayed different levels of high production and up-regulation, 10 lipoproteins were significantly up-regulated during the exponential phase and only SAR0230 lipoprotein had significant down-expression level in this growth phase, whilst 14 lipoproteins were significantly up-regulated in the stationary phase. SAR0206 was among the lipoproteins that showed a decreased quantity during the stationary phase also this protein had no expression in early exponential phase. SAR2457, SAR1881, SAR1288, SAR0872 and SAR0761 showed a constant level of production during all three times of bacterial growth, these 4 lipoproteins displayed up to 6 fold up-regulation over the times of experiment.

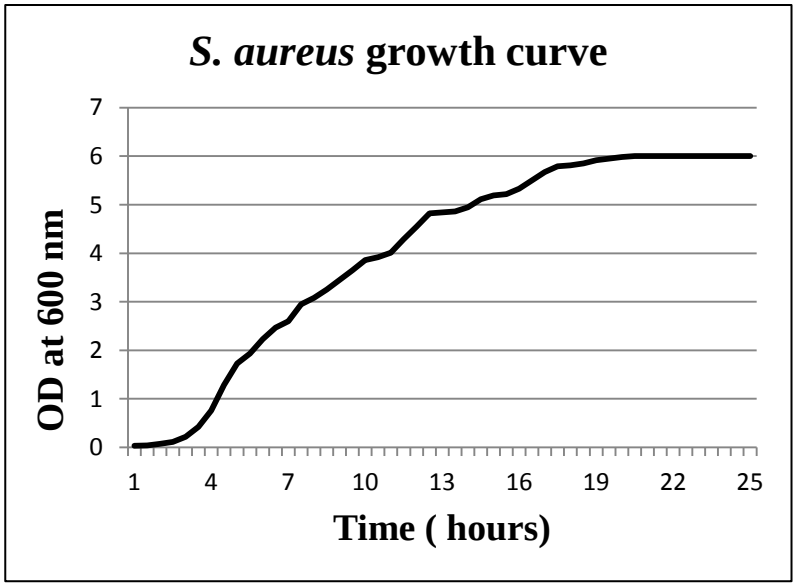

Figure 15. The growth curves of *S. aureus* MRSA252 at 37°C in Tryptone soya broth
The growth curve was plotted following the OD600 measured every 30 min.

Table 17. Comparison of the lipoprotein expression in *S. aureus* MRSA252

List of lipoprotein detected by LC-MS/MS in different bacterial culture conditions (6 h, 15 h and 24 h), label-free relative protein quantification were calculated to show the fold changes levels of up and down regulation, with a fold changes ratio above and below the threshold of ±2.

Annotation	UniProtKB accession	ROF	6 h	15 h	24 h	up/down regulation expression ratio		
						15/6 h	24/6 h	24/15 h
Extracellular solute-binding lipoprotein	Q6GK77	*SAR0230*	2.52E+09	7.50E+08	9.61E+08	**-2.98**	-0.38	1.28
Putative lipoprotein	Q6GIW5	*SAR0730*	3.52E+09	3.25E+09	8.76E+09	-0.92	**2.49**	**2.70**
Putative ferrichrome-binding lipoprotein	Q6GEF0	*SAR2368*	1.31E+11	1.20E+11	2.00E+11	-0.91	1.54	1.67
Efem/EfeO family lipoprotein	Q6GJX5	*SAR0340*	6.39E+08	4.81E+08	8.36E+08	-0.75	1.31	1.74
Putative lipoprotein	Q6GE23	*SAR2500*	8.06E+09	6.04E+09	2.55E+10	-0.75	**3.16**	**4.22**
Putative lipoprotein	Q6GJS0	*SAR0396*	7.58E+10	4.92E+10	4.02E+10	-0.64	-0.53	-0.81
Putative lipoprotein	Q6GGG8	*SAR1608*	4.17E+10	2.57E+10	1.44E+10	-0.61	-0.34	-0.56
Putative lipoprotein	Q6GFF4	*SAR1995*	5.81E+10	3.12E+10	3.71E+10	-0.53	-0.64	1.19
Transport system binding lipoprotein	Q6GEQ2	*SAR2268*	3.42E+10	3.54E+10	6.15E+10	1.04	1.80	1.74
Extracellular solute-binding lipoprotein	Q6GE19	*SAR2504*	6.94E+11	7.42E+11	5.26E+11	1.07	-0.75	-0.7
Putative lipoprotein	Q6GJS7	*SAR0390*	2.18E+11	2.56E+11	5.11E+11	1.17	**2.34**	**2**
Putative lipoprotein	Q6GJM8	*SAR0445*	3.22E+09	3.94E+09	4.78E+09	1.22	1.48	1.21
Putative lipoprotein	Q6GJN3	*SAR0438*	5.46E+08	6.91E+08	1.97E+09	1.27	**3.61**	**2.85**
Putative lipoprotein	Q6GKD3	*SAR0174*	1.23E+09	1.60E+09	1.93E+08	1.30	-0.15	-0.12
Putative lipoprotein	Q6GHM2	*SAR1189*	4.17E+09	5.61E+09	5.41E+09	1.35	1.3	-0.96
Lipoprotein	Q6GIH7	*SAR0872*	1.28E+11	1.76E+11	5.47E+11	1.39	**4.31**	**3.11**
Putative solute-binding lipoprotein	Q6GE27	*SAR2496*	3.97E+10	5.68E+10	1.48E+10	1.43	-0.37	-0.26
Putative lipoprotein	Q6GGS6	*SAR1494*	2.55E+08	3.73E+08	8.55E+08	1.46	**3.36**	**2.29**

Table 17-continued

Annotation	UniProtKB Accession	ROF	6 h	15 h	24 h	Up & down regulation expression ratio		
						15/6 h	24/6 h	24/15 h
Putative lipoprotein	Q6GJN0	*SAR0443*	3.39E+08	5.15E+08	3.27E+08	1.52	-0.96	-0.63
ABC extracellular binding	Q6GJ39	*SAR0641*	2.72E+11	4.11E+11	8.95E+11	1.52	**3.31**	**2.18**
Putative lipoprotein	Q6GGL2	*SAR1558*	6.28E+09	9.80E+09	1.75E+10	1.56	**2.78**	1.78
Putative lipoprotein	Q6GKG2	*SAR0145*	3.07E+09	4.80E+09	1.58E+10	1.57	**5.11**	**3.27**
Putative transport system lipoprotein	Q6GJ59	*SAR0618*	3.16E+10	5.49E+10	4.93E+09	1.74	-0.15	-0.89
Putative lipoprotein	Q6GIQ1	*SAR0794*	1.22E+09	2.32E+09	5.41E+08	1.91	-0.44	-0.23
Putative lipoprotein	Q6GE24	*SAR2499*	9.33E+09	1.79E+10	1.92E+09	1.92	-0.21	-0.10
Lipoprotein	Q6GJL0	*SAR0463*	4.73E+10	9.12E+10	2.02E+11	1.93	**4.27**	**2.21**
Putative lipoprotein	Q6GFQ6	*SAR1881*	5.80E+08	1.19E+09	2.45E+09	**2.05**	**4.22**	**2.06**
Putative lipoprotein	Q6GDX6	*SAR2546*	4.87E+09	1.02E+10	2.00E+09	**2.11**	-0.41	-0.19
Putative lipoprotein	Q6GK91	*SAR0216*	2.11E+09	4.78E+09	4.46E+08	**2.27**	-0.21	-0.93
Putative lipoprotein	Q6GHC4	*SAR1288*	1.32E+10	3.10E+10	7.51E+10	**2.34**	**5.71**	**2.43**
Putative lipoprotein	Q6GHZ3	*SAR1066*	8.67E+09	2.06E+10	2.16E+09	**2.37**	-0.24	-0.11
Putative lipoprotein	Q6GE63	*SAR2457*	5.92E+10	1.43E+11	3.22E+11	**2.43**	**5.45**	**2.25**
Putative lipoprotein	Q6GIK7	*SAR0839*	1.40E+10	3.41E+10	2.61E+10	**2.44**	1.86	-0.76
Putative lipoprotein	Q6GKK2	*SAR0106*	8.05E+08	1.99E+09	8.81E+08	**2.47**	1.10	-0.44
Putative lipoprotein	Q6GIT4	*SAR0761*	7.76E+09	2.39E+10	5.13E+10	**3.07**	**6.61**	**2.15**
Transport extracellular binding protein	Q6GI97	*SAR0953*	9.60E+07	5.18E+08	1.21E+08	**5.41**	1.26	-0.23
Putative lipoprotein	Q6GDC3	*SAR2763*	0.00E+00	1.75E+08	0.00E+00	------	------	------
Extracellular sugar-binding lipoprotein	Q6GKA1	*SAR0206*	0.00E+00	6.10E+08	1.37E+08	------	------	-0.22

4.3 Discussion

Few studies have investigated the characterisation of *S. aureus* lipoproteins and their biological and pathological roles. *S. aureus* genome codes ~ 2700 proteins in total with almost half of them with unknown or predicted functions. *S. aureus* genome has at least 50 genes that harbour the type II signal sequence for lipoprotein and approx. 35 of them are associated with a known or predicted function. Also, some of them are employed in ABC transporters and in acquiring nutrients from the environment (Stoll *et al*., 2005). However just because a gene is present it is not necessarily expressed. In order to visualise the complete proteome of *S. aureus* lipoproteins, one of the main objectives of this study was to perform a highly comprehensive proteomic analysis of *S. aureus* lipoprotein, first step was employing gel-based proteomics techniques by using classical 1D and 2-DE protein separation complemented by LC-MS/MS method as it has been extensively used in several studies. Secondly, the gel-free approach of large scale proteome quantification by in-solution proteins digestion and LC-MS/MS analysis was used. In order to analyse lipoproteins of *S. aureus* the SDS-PAGE gel based method combined with LC-MS/MS was used to determine the lipoprotein quantitation. This has showed a low sensitivity to detect the low abundance lipoproteins as only the highly abundant lipoproteins but the majority of lipoproteins could not be detected after in-gel trypsin digestion. There were only 6 lipoproteins that were detected at the extraction at late exponential phase of bacterial growth.

The failure and low producibility of SDS-PAGE efficiency may due to different reasons including the manual selection of protein band sizes that has a considerable disadvantage which makes it difficult to differentiate between the band sizes. In addition, many reports have reported some of the SDS-PAGE gel limitations including the detergents used in cell solubilisation which can give rise to problems during enzymatic digestion and in the subsequent LC-MS analysis (Hustoft *et al.*, 2012). Detergents such as sodium dodecyl sulfate (SDS) can inhibit the peptide digestion and dominate mass spectra according to their ionizability and great abundance compared to target peptides (Yeung *et al.*, 2008). The six lipoproteins detected by SDS-PAGE were the most abundant lipoproteins, which was confirmed by their abundance in the large scale protein quantification by LC-MS/MS. The reason for this was not clear as the 2-DE gel approach is usually subject to some restrictions of detecting proteins of high molecular weights proteins and pI values. The

method is also limited dynamic range of IPGs strips and handling of hydrophobic proteins (Rabilloud *et al.*, 2009). 2-DE protein spots patterns derived from all samples of different sizes immobilized pH gradients (IPGs) strips were not similar to each other due to the high abundance of some contaminating non-lipoproteins. During in-gel trypsin digestion preparation prior to LC-MS analysis, the tryptic peptide digests were not treated with Zip-Tips to concentrate and purify the samples for further analysis. In the 2-DE method the intensity of protein spots stain was used to determinate the quantity of an individual protein, this quantification technique were inaccurate and unreliable to determine the protein abundance in examined samples, the differences between the protein spots were a result of the failure of the software to identify the spots or were not detected when the unexperienced user during the gel quantification. Spots on 2D gels often contain more than one protein (Kondo, 2008). Becher *et al*. also demonstrated that 2-D gels cover only 25% of the *S. aureus* cytosolic proteome (Becher *et al.*, 2009a). However, due to the poor ability of both 1D and 2D gel electrophoresis to cover the all targeted lipoproteins, these approaches were replaced by a higher resolution and more powerful gel-free quantitative proteome LC-MS/MS methods for better protein quantitation. In addition, the MALDI/TOF machine that had been used in the beginning of this study was out of order for a long time for these reasons the electrophoretic separations by 2-DE approach was terminated.

In recent years there has been more interest to use the label-free approaches for large scale protein mixtures quantitative analysis by LC-MS/MS, in this procedure samples are compared to each other after independent analyses to show a high proteome coverage results. This technique avoids the weaknesses of gel electrophoresis. In-solution protein digestion is a single step method in a single buffer which is more automatable and reduces sample handling during the whole procedure, also to avoid the difficulties during the process of peptide recovery from gels. The main challenge in investigating bacterial proteome is the recoveries of a sufficient amount of proteins from bacterial cells, as different proteins are variable in concentration during the growth phases. Protein abundance usually correlates with the gene expression level, but this correlation between protein and mRNA abundance may differ due to different biological changes of translation, protein degradation and many post-translational modifications (Greenbaum *et al.*, 2002). Label-free approaches are the most accurate mass spectrometric techniques that provide higher dynamic range of proteins quantification (Neilson *et al.*, 2011).

Analytical study to investigate the relationship between transcription, translation and cellular protein turnover in the yeast *Saccharomyces cerevisiae* by comparing protein to mRNA ratios and the translational activity revealed a clear correlation between protein abundance and mRNA are significantly related to biological functional e.g. cellular metabolism and energy (Beyer *et al.*, 2004).

Different expression patterns were observed for the 38 lipoproteins of *S. aureus* were detected. The high accuracy of label-free approach showed the lipoproteins were produced in variable amount during the course of experiment. The undefined functions for most of these lipoproteins it was difficult to find out a reasonable answer for the changeable expression levels. Lipoproteins responsible for iron uptake were intensely over-expressed under iron limitation conditions *in vitro* and *in vivo* suggested their importance in bacterial growth and survival (Allard *et al.*, 2006). In this experiment of non-infection conditions the majorities of these lipoproteins were not significantly induced. This observation can conclude the function of iron uptake by lipoproteins may improve under the nutrients and iron restriction conditions. Up-regulated expression of iron regulation genes in low iron conditions *in vitro* and *in vivo* signifying their important role in iron acquisition (Allard *et al.*, 2006).

Comprehensive gel-free and LC-MS/MS-based quantitative proteome relative quantitative analyses of *S. aureus* COL relative quantitative analyses of surface-associated proteins under infection-relevant situation using three complementing approaches (trypsin shaving, biotinylation and precipitation of the supernatant approaches) identified more than 75% of expressed lipoproteins, both biotinylation and trypsin shaving methods identified 42 lipoproteins, including 23 lipoproteins were identified by biotinylation, whereas precipitation of the supernatant approach was able to identified 18 lipoproteins, these methods were successful for the identification of cell surface and secreted lipoprotein fraction of *S. aureus* (Hempel *et al.*, 2011).

Chapter five

Transcription of *S. aureus* lipoprotein genes studied using quantitative RT-PCR, RNA sequencing and during pathogenesis in *C. elegans*

5. Transcription of *S. aureus* lipoprotein genes studied using quantitative RT-PCR, RNA sequencing and during pathogenesis in *C. elegans*

5.1 Introduction

Little is known about the expression of lipoprotein genes in *S. aureus* during infection conditions in human or any mammalian hosts. Several studies on *S. aureus* gene expression on *in vitro* and *in vivo* conditions including transcriptome analysis during infection have revealed many features of *S. aureus* virulence and pathogenesis. The transcriptome has provided powerful and effective tools for the comprehensive investigation of gene expression profiles and transcriptional levels in bacterial cells (Gilbert *et al.*, 2008).

Quantitative real-time PCR is a useful technique used in both research and diagnostic fields, which provide an absolute copy number of particular target genes (Mullis, 1990). RNA sequencing is a high throughput tool that has been used for transcriptomic research to provide whole genome coverage and is able to analyse all of the expressed transcripts of the tested organism (Oshlack *et al.*, 2010). Both approaches were used in this study to identify and evaluate the expression of the *S. aureus* lipoprotein genes.

Initially, the simplicity of invertebrates as animal models of infection present them as a perfect choice to overcome the logistical limitations of animal models. The nematode *C. elegans* has proved an excellent and useful model of *S. aureus* infection assays to identify host innate immune response and for the screening of pathogen/host-specific pathogenicity factors (Irazoqui *et al.*, 2010a, JebaMercy and Balamurugan, 2012). The percentage of human genes found to be homologous to *C. elegans* genes was 55%, this high homology makes *C. elegans* a good model to discover the roles of some human proteins (Lodish *et al.*, 2000). *C. elegans* has no cell-mediated immunity system, while it has a set of complex innate immune mechanisms in their response against infections including the avoidance behaviours (Pradel *et al.*, 2007, Pujol *et al.*, 2001). The immunity system of *C. elegans* mainly depends on the secretion and action of antimicrobial molecules based on production of antimicrobial proteins as a part of the immune response against bacterial infections (Wong *et al.*, 2007). The accumulation of *S. aureus* within the nematodes digestive tracts kills the *C. elegans* over the course of few days (Garsin *et al.*, 2001,

Irazoqui *et al.*, 2010a, Sifri *et al.*, 2003). One of the objectives of this part of study was to evaluate the lipoprotein genes expression levels during different growth phases by qRT PCR, the analysis was carried out for some selected genes and transcript data were expressed as fold change values.

In order to gain a better understanding of the mechanisms used by *S. aureus* to survive in host and cause diseases, the transcriptome especially for lipoprotein genes in *S. aureus* was studied during infection and non-infection conditions. RNA sequencing analysis performed by a colleague provides an opportunity to study lipoprotein genes expression of *S. aureus* 8325-4 strain. Secondly; RNA was isolated from *S. aureus* MRSA252 during infection of *C. elegans* and subjected to RNA-seq transcriptome analyses to identify the expression of lipoprotein genes levels within an infected host after establishment of infection and to obtain a comprehensive view for the transcriptional response of *S. aureus* during host infection. Additionally, in this study the expression of *C. elegans* genes was also investigated to show the global transcriptional changes for up-and down-regulation of host genes including innate immune responses.

5.2 Results

5.2.1 Detection of lipoprotein genes expression level of *S. aureus* strains *in-vitro* by qRT-PCR

In order to investigate whether lipoproteins were differentially expressed in *S. aureus* at different levels during the growth phases, expression patterns for 5 lipoprotein genes of different sizes were randomly selected and examined by quantitative reverse-transcriptase PCR (qRT-PCR) for three *S. aureus* strains (MRSA252, RN4282 and T1) and expression levels of each gene between the examined *S. aureus* strains were determined as shown in figures 16, 17 and table 18. The results showed a different pattern of up-or-down regulation for each gene within different stages of growth. Gene expression levels were calculated by both Pfaffl and Livak methods and gave the same results, MRSA252 at 4 h incubation time (early exponential phase) was chosen as calibrator. *SAR0444* gene in *S. aureus* MRSA252 and T1 showed a similar expression level at both time-points of measurement, both strains had a slight up-regulation level for *SAR0444* gene after 8 h (exponential phase), while *SAR0444* gene in RN4282 strain had a constant expression level with 1 fold level up-regulation during the experiment time-point of 4 and 8h. *SAR0216* gene had a different expression level among RN4282 strain to show expression level at both time-points with ~1 fold up-regulation, while MRSA252 and T1 strains had parallel expression levels during the growth phases. *SAR0390* lipoprotein gene was the most stable gene with more similar expression level in tested strains in examined growth stages, but slight variance in the expression level was observed in RN4282 and T1 strains at 8 h time-point. *SAR0340* lipoprotein transcript showed a equal level of expression in the tested strains in different growth stages, this gene had a slight higher expression level in the examined strains at 8 h. *SAR0730* gene expression patterns among the 3 tested strains had almost identical expression levels in the different growth phases and the expression levels were similar at 4 and 8 h in all strains. Figure 15 in section 4.2.3 are shown the growth curves of *S. aureus* MRSA252 culture at 37°C in Tryptone soya broth used in the experiments.

Table 18. Expression fold level of 5 lipoprotein genes in 3 *S. aureus* strains examined by quantitative real-time PCR

Results were calculated with two different methods, 2−ΔΔCT (Livak) method and the ΔCT method, *S. aureus* MRSA252 at 4 h incubation time was chosen as calibrator. The two methods been used have given the same results.

Gene ORF	Predicted function	Gene expression levels of different *S. aureus* strains				
		MRSA252 8h	RN4282 4h	RN4282 8h	T1 4h	T1 8h
SAR0444	putative lipoprotein	0.1	1.16	1	0	0.1
SAR0216	putative lipoprotein	0.1	0.81	1.02	0	0.35
SAR0390	putative lipoprotein	0.15	0	0.33	0.1	0.20
SAR0340	putative lipoprotein	0.55	-0.2	0	-0.1	0.51
SAR0730	putative lipoprotein	0.36	0.39	0.60	0	0.5

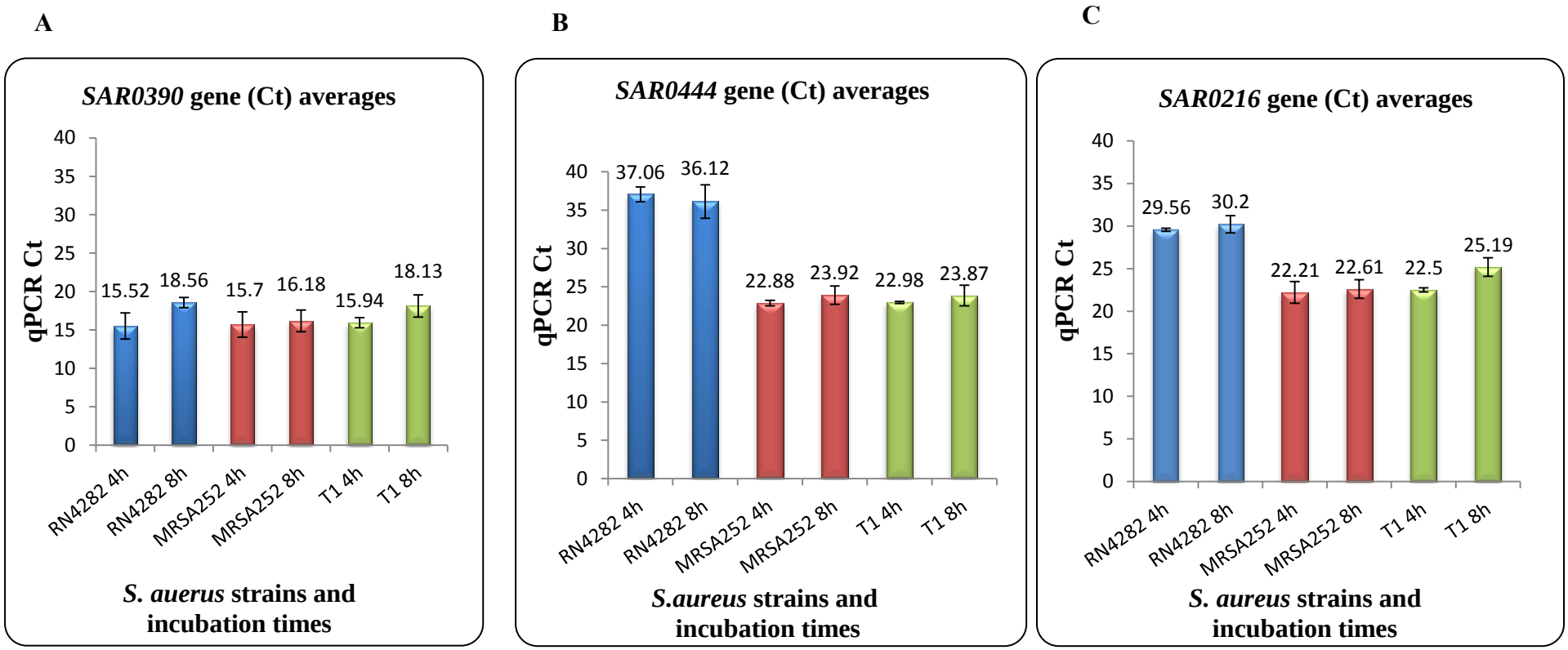

Figure 16. Comparisons of differential expression fold patterns of lipoprotein genes as determined by real-time quantitative PCR assay

Three *S. aureus* strains (RN4282, MRSA252 and T1) were cultivated separately in the same conditions and RNA was extracted in different incubation times (4 and 8 h), **A**: qPCR fold change of *SAR0390* gene, **B**: qPCR fold change of *SAR0444* gene, **C**: qPCR fold change of *SAR0216*gene, Ct values (triplicate) for each lipoprotein gene were plotted against each strain and the incubation times. The relative expression levels of selected genes were calculated using both Pfaffl and Livak methods. Error bars represent standard deviation of mean expression values.

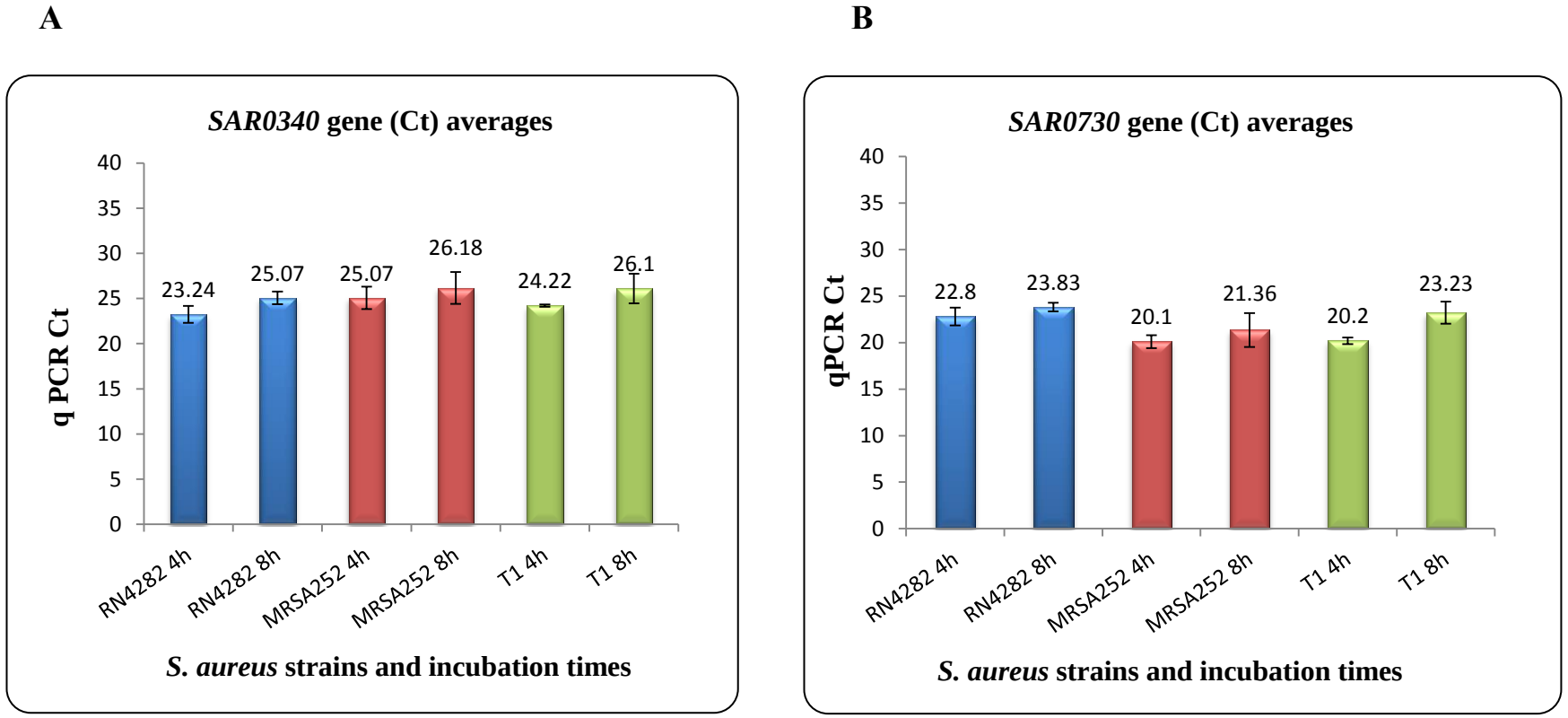

Figure 17. Comparisons of differential expression patterns of lipoprotein genes as determined by real-time quantitative PCR assay

Three *S. aureus* strains (RN4282, MRSA252 and T1) were cultivated separately in the same conditions then RNA was extracted in different incubation times (4 and 8 h), **A**: qPCR fold change of *SAR0340* gene, **B**: qPCR fold change of *SAR0730* gene, Ct values (triplicate) for each lipoprotein gene were plotted against each strain and the incubation times. The relative expression levels of selected genes were calculated using both Pfaffl and Livak methods. Error bars represent standard deviation of mean expression value.

5.2.2 Transcriptome comparison of *S. aureus* 8325-4 lipoprotein genes

The first experiment to compare the transcriptome of lipoprotein genes in *S. aureus* 8325-4 was to compare gene expression at three growth time-points (1, 12 and 24 h), bacteria were inoculated into Iso-sensitest broth then the RNA was extracted. These time-points of grown culture have been chosen as a control samples in other expermints of investigation the effect of mupirocin on *S. aureus* 8325-4. The RNA was subjected to RNA-seq and the data analysis was performed to examine the differences between samples groups in order to evaluate the differential expression of genes at the three times. The results showed that 54 lipoprotein genes were detected and transcriptome comparisons to evaluate the expression differences are summarized in table 19. Transcription of 54 genes showed that genes were up and down regulated differently during the course of experiment. The comparison of gene expression between the 1 and 12 h growth showed that no lipoprotein genes were down-regulated and 20 genes were significantly up-regulated, while the ratio comparison between 1 h and the stationary phase at 24 h showed that 35 lipoprotein genes were down-regulated and 14 genes were up-regulated. Additionally the ratio of transcription of the genes changed between the exponential (12 h) and stationary phase (24 h) showed 5 genes were significant down-regulated and 4 genes were significant up-regulated.

The transcriptome compression of lipoprotein genes in *S. aureus* 8325-4 strain under non-infection condition and the transcripts comparison for three growth time-points (1, 12 and 24 h), the results obtained showed that 23 lipoprotein transcripts analogues for *S. aureus* MRSA252 including *SAR1189, SAR0953, SAR0438, SAR0439, SAR2363, SAR1495, SAR2536, SAR1494, SAR2554, SAR0444, SAR0390, SAR0872* and *SAR1066* were significantly up-regulated with 2 to 13 fold in different patterns.

Table 19. Transcriptome comparison of lipoprotein genes in *S. aureus* 8325-4 at different phases of growth

Three growth time-points (1, 12 and 24 h) were compared and the ratios of transcripts up and down-regulation between the times, determined with a fold change ratio above and below the threshold of ±2, significantly differentially expressed genes were defined with FDR-adjusted P-value < 5%.

Homologous *S. aureus* MRSA252 ORF	*S. aureus* 8325-4 locus tag	Reads per kilobase of transcript per million			Ratio 12h/1h	Ratio 24h/12h	Ratio 24h/1h
		1h	12h	24h			
SAR1402	SAOUHSC_01389	4081	53630	1659	**13.14**	**-32.32**	**-2.45**
SAR1189	SAOUHSC_01180	1430	14700	1740	**10.3**	-0.12	1.22
SAR0839	SAOUHSC_00808	1139	10191	956	**8.94**	**-10.66**	-1.19
SAR0953	SAOUHSC_00927	4300	26600	2190	**6.19**	-0.08	-0.51
SAR0145	SAOUHSC_00105	265	1454	526	**5.47**	**2.76**	-0.50
SAR1878	SAOUHSC_01917	1171	6241	1025	**5.32**	**-6.08**	-1.14
SAR0438	SAOUHSC_00405	1690	8980	2130	**5.31**	-0.24	1.26
SAR0438	SAOUHSC_00405	1688	8980	2128	**5.31**	**4.21**	0.79
SAR0439	SAOUHSC_00402	704	2380	819	**3.38**	-0.34	1.16
SAR0443	SAOUHSC_00402	704	2382	819	**3.38**	**2.90**	0.85
SAR2363	SAOUHSC_02549	9040	29000	4520	**3.21**	-0.16	-0.51
SAR0761	SAOUHSC_00717	2604	7835	3191	**3.00**	**-2.45**	0.81
SAR0954	SAOUHSC_00928	373	1059	805	**2.83**	-1.31	0.46
SAR1495	SAOUHSC_01512	1820	4840	1770	**2.66**	-0.37	-0.97
SAR2536	SAOUHSC_02742	1180	3030	1940	**2.57**	-0.64	1.64
SAR1494	SAOUHSC_01508	2180	4980	3170	**2.28**	-0.64	1.45
SAR2763	SAOUHSC_03016	1092	2450	2821	**2.24**	0.86	0.38
SAR2554	SAOUHSC_02767	1990	4280	1500	**2.15**	-0.35	-0.75
SAR0444	SAOUHSC_00404	1500	3180	1360	**2.12**	-0.43	-0.91
SAR0872	SAOUHSC_00844	2490	5050	932	**2.03**	-0.19	-0.37
SAR0201	SAOUHSC_00170	873	1700	1360	1.95	-0.8	1.56
SAR0106	SAOUHSC_00054	2051	3980	1438	1.94	**2.76**	-1.42
SAR0390	SAOUHSC_00356	1260	2420	2620	1.92	1.08	**2.08**
SAR0340	SAOUHSC_00325	565	1060	1040	1.88	-0.98	1.84
SAR0641	SAOUHSC_00634	3330	6050	2650	1.82	-0.44	-0.8
SAR0463	SAOUHSC_00426	6060	10900	2420	1.80	-0.22	-0.4
SAR0230	SAOUHSC_00201	3030	5330	2360	1.76	-0.44	-0.78
SAR2268	SAOUHSC_02430	5100	8790	777	1.72	-0.09	-0.15
SAR1011	SAOUHSC_00976	1690	2890	1910	1.71	-0.66	1.13
SAR1106	SAOUHSC_01084	462	702	431	1.52	-0.61	-0.93
SAR1879	SAOUHSC_01918	10400	14800	2410	1.42	-0.16	-0.23
SAR0794	SAOUHSC_00754	4140	5740	3520	1.39	-0.61	-0.85
SAR0174	SAOUHSC_00137	171	236	218	1.38	-0.92	1.27
SAR2457	SAOUHSC_02650	5600	7410	2570	1.32	-0.35	-0.46
SAR0118	SAOUHSC_00074	821	1040	360	1.27	-0.35	-0.44
SAR1881	SAOUHSC_01920	1490	1870	1440	1.26	-0.77	-0.97
SAR2573	SAOUHSC_02789	873	1067	395	1.22	**- 2.70**	**-2.21**
SAR2504	SAOUHSC_02698	7390	8600	7160	1.16	-0.83	-0.97

Table 19-continued

Homologous *S. aureus* MRSA252 ORF	*S. aureus* 8325-4 locus tag	Reads per kilobase of transcript per million			Ratio 12h/1h	Ratio 24h/12h	Ratio 24h/1h
		1h	12h	24h			
SAR0730	SAOUHSC_00685	7810	8660	7010	1.11	-0.81	-0.9
SAR2496	SAOUHSC_02690	13600	13900	4340	1.02	-0.31	-0.32
SAR2499	SAOUHSC_02694	24800	8180	2560	-0.33	-0.31	-0.1
SAR0206	SAOUHSC_00176	1850	664	814	-0.36	1.23	-0.44
SAR1066	SAOUHSC_01039	32500	13500	101000	-0.42	**7.48**	**3.11**
SAR0790	SAOUHSC_00749	2460	1240	816	-0.51	-0.66	-0.33
SAR2546	SAOUHSC_02759	28600	17000	2340	-0.59	-0.14	-0.08
SAR2500	SAOUHSC_02695	2670	1600	418	-0.62	-0.26	-0.16
SAR1034	SAOUHSC_01002	374000	286000	28600	-0.77	-0.11	-0.08
SAR0396	SAOUHSC_00362	4700	3830	3010	-0.82	-0.79	-0.64
SAR0618	SAOUHSC_00613	4470	3720	857	-0.83	-0.23	-0.19
SAR2368	SAOUHSC_02554	49200	41200	6860	-0.84	-0.17	-0.14
SAR0216	SAOUHSC_00185	3890	3400	1500	-0.87	-0.44	-0.39
SAR2179	SAOUHSC_02327	32300	30000	28900	-0.93	-0.96	-0.9
SAR1995	SAOUHSC_02121	30900	29800	17300	-0.96	-0.58	-0.56
SAR1608	SAOUHSC_01627	2550	2520	2950	-0.99	1.17	1.16

5.2.3 Changes in gene expression during infection using *C. elegans* as a model

C. elegans killing assay was performed using *S. aureus* MRSA252. Nematodes were incubated at 20°C on TS agar and the nematodes were observed and scored as alive or dead on a daily basis. The killing assay results revealed that exposure to *S. aureus* led to a significant decrease in life span with ~ 2 d required for half of the nematodes to die (figure 18) compared with nematodes grown on non-pathogenic *E. coli* OP50 which had a mean life span of ~ 17 d (figure 19). *S. aureus* infected nematodes were examined under a Nomarski differential interference contrast microscope showed that the nematodes incubated with MRSA252 were smaller in size as a result of starvation unlike the non-infected worms fed on *E. coli* OP50. Nematodes displayed a distended intestinal tract that was presumable colonized with *S. aureus*; this was particularly evident in the anterior gut. In contrast, nematodes fed on *E. coli* had normal intestinal lumens (figure 21 A and B). The newly matured self-fertilizing hermaphrodite nematodes can produce a few hundred of progeny in the normal conditions but under the influence of infection this process was affected and led to the accumulation of fertilized eggs inside the body (figure 20 A), resulting in the appearance of the "bag-of-worms" phenotype.

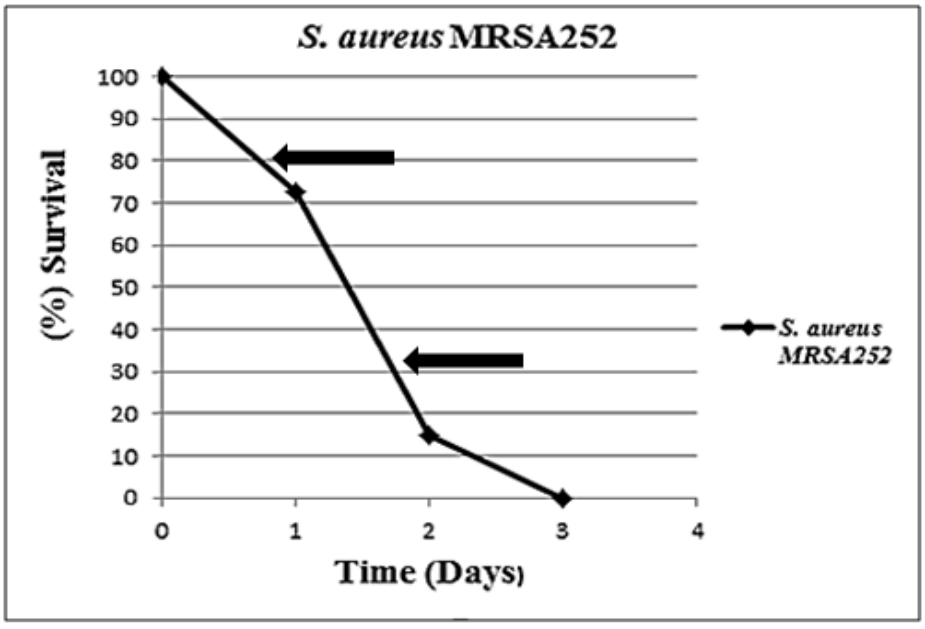

Figure 18. *C. elegans* survival in TS agar at 20°C with *S. aureus* MRSA252

Kaplan-Meier survival curve of N2 Bristol wild-type (WT) *C. elegans*, the graph shows survival curve of *C. elegans* cultivated in *S. aureus* MRSA252. The arrows to show the sampling times (16 and 40 h).

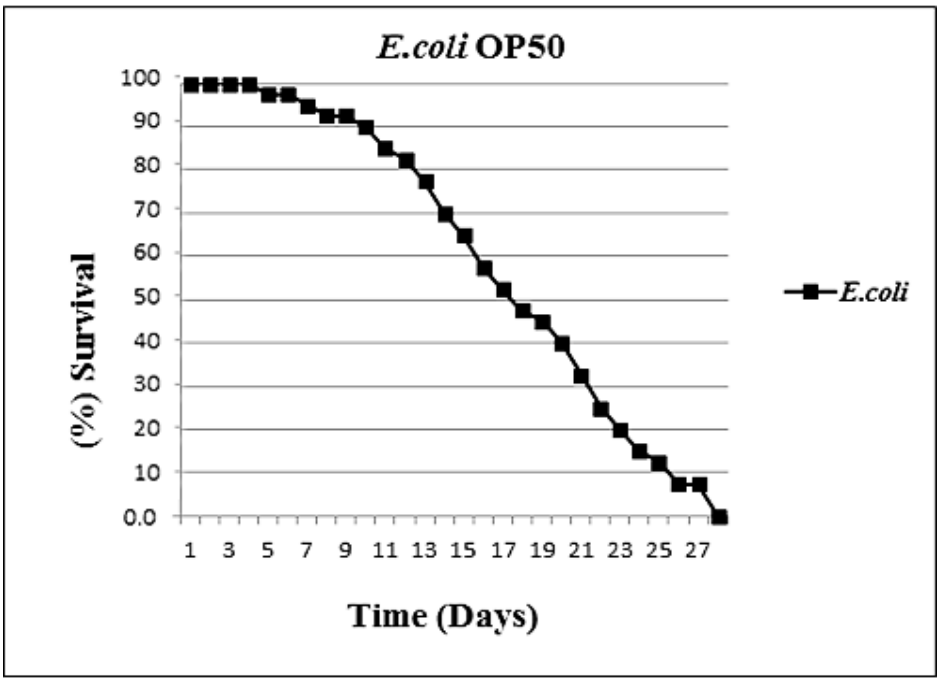

Figure 19. N2 Bristol wild-type (WT) *C. elegans* survival in NGM agar at 20°C with *E. coli* OP50

The graph shows survival curve of *C. elegans* cultivated in non-pathogenic *E. coli* OP50, Kaplan-Meier survival curve of *C. elegans* fed on *E. coli* OP50 was used as a control.

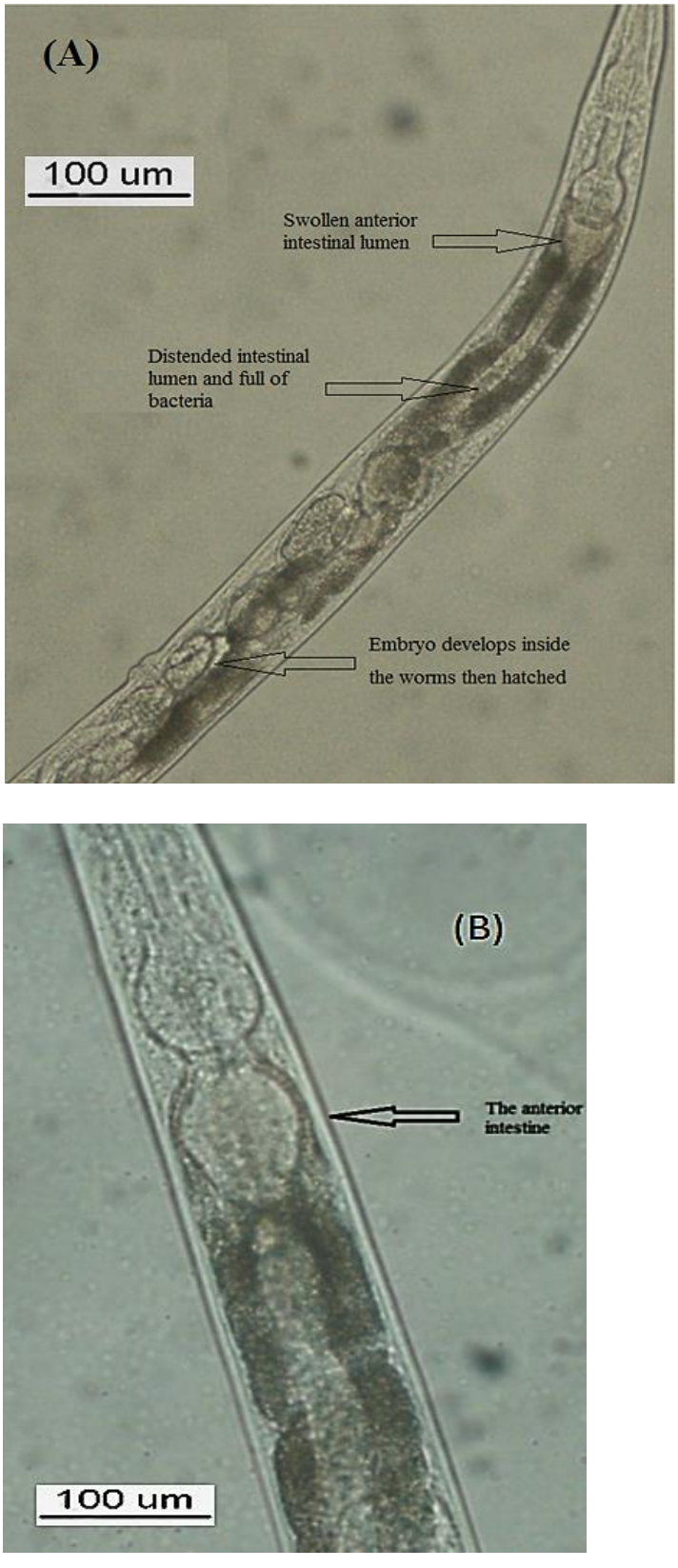

Figure 20. Images of *C. elegans* fed on *S. aureus*
The cellular structures were visualized under Nomarski differential interference contrast (DIC) microscope after 24 h of feeding are shown in images A and B, *S. aureus* colonizes the entire intestine of the worms and killed them within a few days.

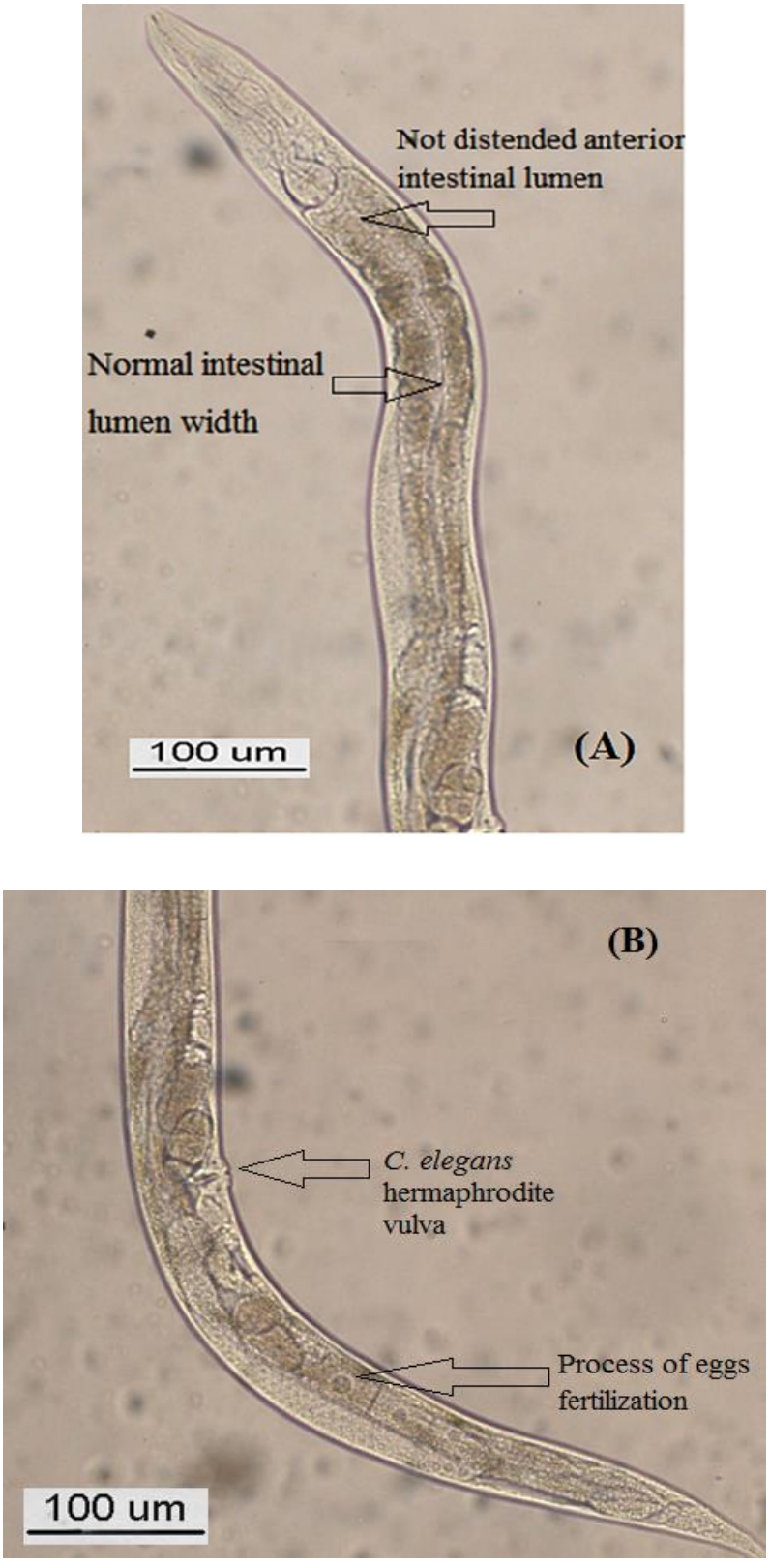

Figure 21. Images of healthy *C. elegans* fed on *E. coli* OP50

The cellular structures were visualized under Nomarski differential interference contrast (DIC) microscope after 24 h of feeding are shown in images A and B, the standard laboratory food *E. coli* OP50 of *C. elegans* does not colonize and cause any mortality.

5.2.4 Transcriptomics of *S. aureus* MRSA252 lipoprotein genes obtained from *C. elegans* infection model

To show the molecular features of the *S. aureus* response to infection and to profile the transcriptional pattern of *S. aureus* lipoprotein genes expression levels during *C. elegans* infection challenge, RNA sequencing was used to quantify total genes expression and in particular the lipoprotein genes transcripts from *S. aureus* RNA isolated from infected *C. elegans* after different times of infection. The results were obtained from three independent experiments. The common method of heatmap was used to present the gene expression level for a number of comparable samples by assigning different colours to each transcript, thus groups of transcripts of similar or different expression values can be easily visible. A correlation heatmap matrix for RNA-seq samples based on genes that up-regulated were coloured differently than genes that were down-regulated, consequently providing a simultaneous visual representation of genes expression levels among different samples. Also to show the differentially expressed profile in each transcript of *S. aureus* MRSA252 during the *C. elegans* infection stages used to comparatively illustrate gene expression levels across a number of different samples as shown in figure 22. Figure 23 shows the correlation heatmap matrix for RNA-seq of *C. elegans* samples. The first step was to compare the levels of lipoprotein genes transcripts of *S. aureus* at each time (16 and 40 h). To evaluate the relative gene expression levels, transcript data were calculated as fold change values for RNA from the samples of *C. elegans* infection model and compared relatively to *S. aureus* MRSA252 transcripts from 16 h *in vitro* culture as a control. Results of the lipoprotein transcriptome are presented in table 20. Tables containing the number of significantly expressed genes were defined as those with FDR-adjusted P-value < 5%, P-values associated with fold change FC were adjusted for multiple testing using the False Discovery Rate (FDR) approach (Benjamini and Hochberg, 1995). Gene expressions levels of individual transcripts were estimated from reads alignment files by using Cufflinks corresponding FPKM values were converted from count numbers and reported in this analysis. The number of fragments per kilobase per million base (FPKM) was used to show the gene expression levels. The table shows how many of these genes were up-regulated and how many of them were down-regulated. The overall transcription of lipoprotein genes from infected samples in the two points of evaluation showed the total number of lipoprotein genes expressed during the time of infection were 62 genes. At the early stage of bacterial infection no gene showed up-regulated transcripts and 8 genes

(13%) were down-regulated (table 20). Three lipoprotein genes expressed at the late stage of infection (5%) were up-regulated and 17 genes (28%) were significantly down-regulated. In contrast, the relative gene expression levels between the two time-points of post-infection showed that 3 lipoprotein genes were up-regulated in this time of evaluation. Lipoprotein genes that were significantly down-regulated at the early and late stages of infection was 6 and 14 lipoprotein genes respectively, they showed significant down expression after the nemato-bacterial infection. The most strongly up-regulated lipoproteins at the two points of 16 and 40 h were 2 and 3 respectively lipoprotein genes as showed ~2 fold change of significant high production. Among the down-regulated set, only two lipoproteins *SAR1878* and *SAR1402* that involved and associated with the inflammatory responses during the early stage of *C. elegans* infection to show >2 fold of significant down-regulation. *SAR2736* and *SAR1565* putative lipoprotein both of them were annotated as (pseudogene) in *S. aureus* transcriptome meta-database (SATMD), but their expression levels during the *C. elegans* challenge course proved their ability to expressed as lipoprotein.

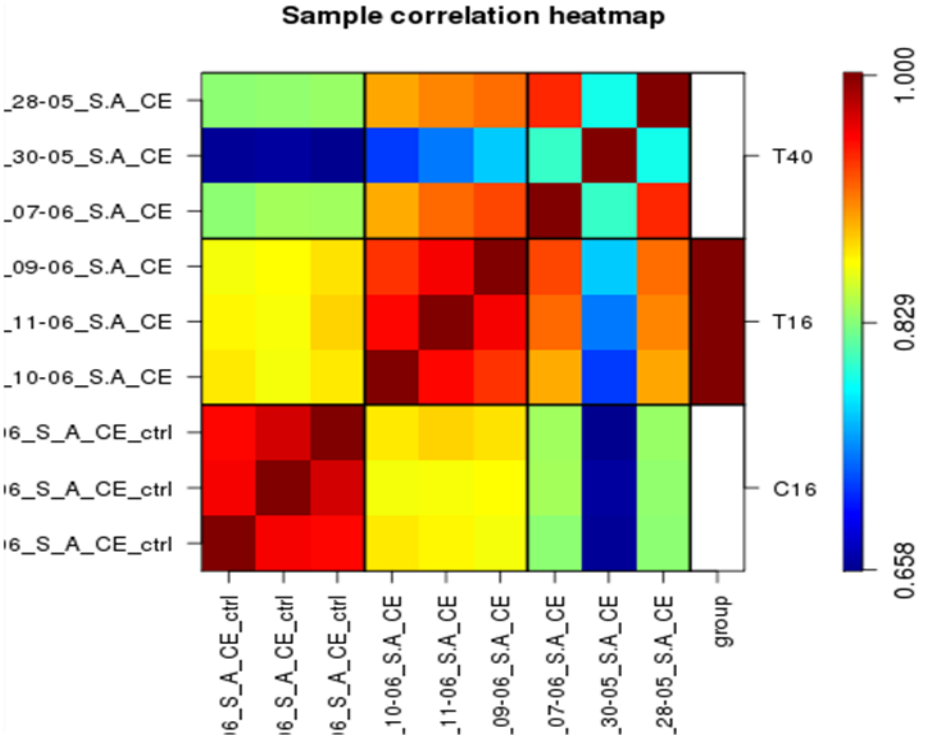

Figure 22. Correlation heatmap matrix for RNA-seq samples
The heatmap based on the genes that up-regulated were coloured differently than genes that were down-regulated to show the differentially expressed profile in each transcript of *S. aureus* MRSA252 during the *C. elegans* infection stages, P-value less than <0.05, colour code values in the heatmap are shown in the side (blue to brown colour scale).

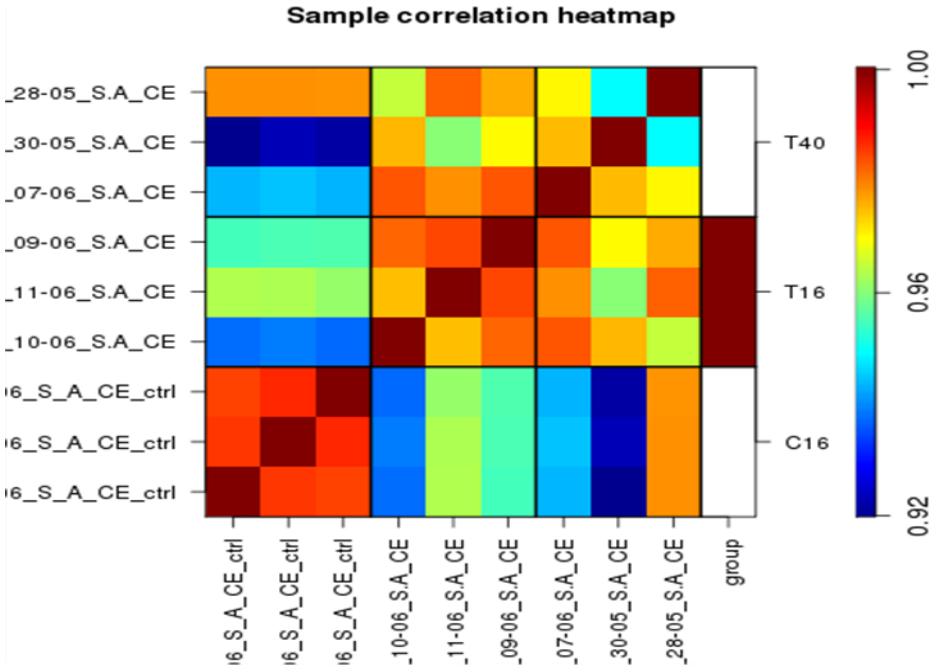

Figure 23. Correlation heatmap matrix for RNA-seq *C. elegans* samples
The heatmap based on the genes that up-regulated were coloured differently than genes that were down-regulated to displaygene expressions at different time points transcript of *C. elegans* RNA-seq sample during the *C. elegans* infection model, P-value less than <0.05, colour code values in the heatmap are shown on the side.

Table 20. Comparison of *S. aureus* MRSA252 lipoprotein genes transcriptome in the *C. elegans* infection model at 16 and 40 h

The transcriptome of two conditions were compared with 16 h *in vitro* control culture to show the ratio fold change, also the two treated samples were compared to show the changes occurred, fold change ratio above and below the threshold of ±2, significantly differentially expressed genes were defined with FDR-adjusted P-value < 5%. ORF's in bold were detected in lipoprotein proteomics experiment.

MRSA252 ORF	Description	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
SAR0106	Putative lipoprotein	3.935536	-0.13	7.99E-01	-1.28	9.14E-02	-1.16	1.43E-01
SAR0118	Lipoprotein	5.362675	0.73	7.47E-02	1.08	2.69E-02	0.35	4.84E-01
SAR0145	Putative lipoprotein	2.864657	-0.91	1.65E-01	-0.29	7.30E-01	0.62	5.08E-01
SAR0174	Putative lipoprotein	2.067345	-0.63	4.25E-01	**-5.98**	1.27E-02	**-5.35**	4.26E-02
SAR0201	RGD-containing lipoprotein	2.228918	0.09	9.19E-01	-1.83	2.05E-01	-1.91	1.93E-01
SAR0206	Putative extracellular sugar-binding lipoprotein	5.81876	-0.75	9.19E-02	**-2.18**	6.27E-04	-1.43	2.92E-02
SAR0216	Putative lipoprotein	3.974462	-1.09	3.82E-02	-1.46	4.41E-02	-0.37	6.36E-01
SAR0230	Putative extracellular solute-binding lipoprotein	5.679551	**-2**	3.18E-06	-1.9	2.70E-04	0.1	8.57E-01
SAR0340	Putative lipoprotein	2.712954	**-3.57**	5.40E-06	**-3.38**	2.52E-03	0.19	9.04E-01
SAR0390	Putative lipoprotein	11.86024	-1.28	1.10E-03	-0.6	1.25E-01	0.68	8.26E-02
SAR0396	Putative lipoprotein	5.674911	-0.1	8.11E-01	0.02	9.71E-01	0.12	8.18E-01
SAR0438	Putative lipoprotein	-3.71531	-1.77	3.34E-01	-1.77	5.83E-01	0	1.00E+00
SAR0439	Putative lipoprotein	0.15784	-1.83	2.20E-01	**-4.76**	1.28E-01	**-2.93**	3.90E-01
SAR0442	Tandem lipoprotein	1.716093	0.62	6.26E-01	1.05	4.82E-01	0.43	7.80E-01
SAR0443	Putative lipoprotein	2.690094	0.51	5.60E-01	0.34	7.60E-01	-0.17	8.82E-01
SAR0444	Putative lipoprotein	3.291042	-0.88	1.09E-01	**-2.36**	1.29E-02	-1.48	1.57E-01
SAR0445	Putative lipoprotein	6.229949	0.77	6.02E-02	0.92	4.77E-02	0.15	7.56E-01
SAR0463	Putative lipoprotein	6.099396	-1.81	1.56E-04	-1.91	2.55E-03	-0.1	8.84E-01

Table 20-continued

MRSA252 ORF	Description	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
SAR0618	transport system lipoprotein	5.46572	-1.37	5.57E-04	**-2.6**	5.57E-06	-1.22	4.60E-02
SAR0641	ABC transporter extracellular binding	10.74544	-0.83	8.38E-02	**-4.58**	2.84E-14	**-3.75**	1.81E-10
SAR0706	Putative protein	-1.16517	0	-	**3.94**	4.25E-02	**3.94**	1.12E-01
SAR0730	Putative lipoprotein	7.813473	-0.25	4.32E-01	0.21	5.60E-01	0.46	2.05E-01
SAR0761	Putative lipoprotein	6.68203	0.74	1.27E-01	0.97	6.49E-02	0.22	6.71E-01
SAR0790	Lipoprotein	6.137203	0.49	3.69E-01	-1.2	7.05E-02	-1.69	1.15E-02
SAR0794	Putative lipoprotein	8.006958	-0.01	9.80E-01	-0.25	4.69E-01	-0.24	4.88E-01
SAR0839	Putative lipoprotein	5.259353	0.16	7.65E-01	-1.43	4.08E-02	-1.59	2.56E-02
SAR0872	Putative lipoprotein	7.612778	0.55	1.81E-01	0.18	7.09E-01	-0.37	4.46E-01
SAR0953	Extracellular binding lipoprotein	9.61191	1.43	1.25E-03	1.46	1.47E-03	0.03	9.45E-01
SAR1011	Transport extracellular binding lipoprotein	5.972712	-0.69	2.90E-01	1.58	3.03E-02	**2.27**	3.34E-03
SAR1034	Putative quinol oxidase polypeptide II precursor	13.41435	-0.98	2.79E-02	-1.4	1.90E-03	-0.42	3.42E-01
SAR1066	Putative lipoprotein	8.718441	-0.58	6.82E-02	-1.15	9.70E-04	-0.57	1.04E-01
SAR1106	Transport extracellular binding lipoprotein	-0.64547	-1.95	2.35E-01	**-4.1**	2.07E-01	**-2.15**	5.24E-01
SAR1189	Putative lipoprotein	5.218548	1.76	1.05E-05	-0.39	5.33E-01	-2.15	2.56E-04
SAR1288	Putative lipoprotein	8.244903	-1.52	1.07E-05	-1.64	2.81E-05	-0.12	7.68E-01
SAR1402	Phosphate-binding lipoprotein	7.665538	**-2.14**	5.93E-04	**2.71**	6.47E-06	**4.85**	1.35E-12

Table 20-continued

MRSA252 ORF	Description	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
SAR1494	Putative lipoprotein	1.53716	-0.58	5.59E-01	-1.43	3.52E-01	-0.85	6.00E-01
SAR1495	Putative lipoprotein	1.1517	-0.09	9.26E-01	0.55	6.74E-01	0.64	6.48E-01
SAR1558	Putative lipoprotein	7.435479	-1.12	2.07E-03	**-2.28**	3.38E-07	-1.16	9.56E-03
SAR1565	Putative lipoprotein (pseudogene)	7.23253	0.4	3.16E-01	0.16	7.10E-01	-0.24	5.90E-01
SAR1608	Putative lipoprotein	6.520394	-0.89	1.27E-02	0.07	8.65E-01	0.96	2.10E-02
SAR1831	beta-lactamase precursor	11.39788	**-2.69**	4.84E-10	**-2.19**	3.86E-07	0.49	2.32E-01
SAR1878	Putative lipoprotein	1.912886	**-2.34**	8.43E-03	-1.16	2.90E-01	1.18	4.01E-01
SAR1879	Putative lipoprotein	11.07714	-1.54	3.07E-03	-1.32	1.13E-02	0.22	6.71E-01
SAR1881	Putative lipoprotein	4.332287	-0.35	5.01E-01	-1.26	7.97E-02	-0.91	2.24E-01
SAR1932	Peptidyl-prolyl cis-isomerase	11.57265	**-3.48**	1.10E-14	**-3.87**	1.71E-16	-0.39	3.60E-01
SAR1995	Putative lipoprotein	8.782224	-0.41	2.05E-01	0.24	4.74E-01	0.65	5.77E-02
SAR2179	Putative membrane protein	8.760973	0.65	5.83E-02	0.69	5.92E-02	0.04	9.24E-01
SAR2268	Putative transport system binding lipoprotein	6.969823	-0.25	5.71E-01	**-2.14**	8.66E-05	-1.89	5.52E-04
SAR2363	Putative molybdate-binding lipoprotein precursor	6.671393	**-2.09**	4.39E-08	**-3.1**	4.16E-10	-1.01	5.07E-02
SAR2368	Putative ferrichrome-binding lipoprotein	7.031337	-0.24	4.78E-01	**-2.62**	7.08E-08	**-2.38**	1.30E-06
SAR2457	Putative lipoprotein	11.99348	1.95	6.63E-07	**2.58**	1.60E-10	0.63	1.01E-01
SAR2470	Putative exported protein	4.535592	-0.28	4.82E-01	-2.02	2.55E-03	-1.74	1.30E-02

Table 20-continued

MRSA252 ORF	Description	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
SAR2496	Putative lipoprotein	6.852125	-1.39	2.02E-04	-1.5	5.94E-04	-0.11	8.00E-01
SAR2499	Putative lipoprotein	8.001797	-1.4	2.41E-04	-0.98	1.93E-02	0.42	3.16E-01
SAR2500	Putative lipoprotein	4.258847	-0.85	1.14E-01	-1.26	7.83E-02	-0.41	5.90E-01
SAR2504	Extracellular solute binding lipoprotein	9.306613	-0.19	5.27E-01	-0.62	6.32E-02	-0.42	2.05E-01
SAR2536	Glycine/carnitine/ binding lipoprotein	6.857105	**-3.67**	9.48E-15	**-3.36**	6.30E-10	0.31	5.90E-01
SAR2546	Putative lipoprotein	7.586737	-1.21	3.55E-04	**-2.99**	2.05E-11	-1.78	7.37E-05
SAR2554	Putative substrate binding lipoprotein	6.289847	-0.28	5.06E-01	-1.83	1.04E-03	-1.55	6.07E-03
SAR2573	Putative lipoprotein	4.205983	-0.45	4.15E-01	-0.73	2.98E-01	-0.29	6.99E-01
SAR2546	Putative lipoprotein	7.586737	-1.21	3.55E-04	**-2.99**	2.05E-11	-1.78	7.37E-05
SAR2736	Putative lipoprotein (pseudogene)	4.018851	1.33	5.66E-02	1.22	1.37E-01	-0.11	8.87E-01
SAR2763	Putative lipoprotein	4.879761	0.79	5.01E-02	-0.01	9.82E-01	-0.8	1.48E-01

5.2.5 Transcriptome analysis of *S. aureus* MRSA252 genes

From the analysis of 2620 genes that were detected in the transcriptome profile of *S. aureus* MRSA252 during the course of *C. elegans* infection, the top one hundred genes up and down-regulated of each infection condition were selected and analysed to show the most induced *S. aureus* genes during infection model. Table 21 summarises the main groups of up-regulated genes of *S. aureus* after 16 h of infection and their predicted function and the top 100 up-regulated genes are shown in table 22. After 16 h of infection the *S. aureus* showed a significant up-regulation expression change in the capsular polysaccharide synthesis enzymes and membrane protein production. Thirty-one of the most affected 100 genes were related to cell wall synthesis proteins, while the predominant produced proteins were enzymes related to biosynthesis of the essential amino acids and enzymes belongs to oxidoreductases (table 22). The major groups of down-regulated genes after 16 h are summarised in table 23 and details are given in table 24. Down-regulated genes during the early infection stage were related to the cell membrane transport systems, and there were a high number of proteins related to metabolism, hypothetical proteins and some transporting ATPase enzymes (table 23). The top down-regulated genes included membrane transporters, enzymes of arginine and ornithine metabolism (table 24).

At the late stage of infection the transcriptome profile showed the most up-regulated genes of *S. aureus* at 40 h were similar to that of early stage of infection including up-regulation of capsular polysaccharide synthesis enzymes, membrane protein synthesis and amino acid transport (tables 25 and 26). The down-regulated genes at 40 h of infection had no clear trends but showed that some hypothetical phage proteins were among the most affected genes, cell wall metabolism protein and exotoxin were also detected in the late stage as shown in tables 27 and 28. The two points of infection (16 and 40 h) were compared to find out the most affected genes as presented in table 29 and 30, among the up-regulated protein were 34 hypothetical protein, 13 membrane protein and several ABC transporter ATP-binding proteins, while the second most up-regulated was *SAR1402* a phosphate-binding lipoprotein which also had a high level of expression during the early infection stage.

Tables 31 and 32 shows the most down-regulated genes at 40 h comparing to 16 h of infection, 11 phage protein, 34 hypothetical proteins and 4 exotoxins were significantly down-expressed in this period between the two stages of infection.

5.2.5.1 *S. aureus* metabolism regulated pathways

Oxygen is one of the most essential growth-limiting factors for most bacterial, both Gram-positive and Gram-negative bacteria regulate their gene expression in response to carbon availability, many bacteria have evolved sophisticated mechanisms to use different carbohydrates to allow a rapid growth; the induction of specific carbohydrate transport and utilization systems in the presence of the carbon source and their repression when a more efficiently utilizable carbohydrate is present, this mechanisms is known as carbon catabolite repression, In many low-GC Gram-positive bacteria carbon catabolite regulation is mediated by the Catabolite control protein A *CcpA.CcpA* regulates more than 100 genes in *S. aureus* (Seidl, Muller et al. 2009). Characteristics of metabolic profiles of *S. aureus* are still unclear and very little information was identified about the affect of different growth conditions on metabolic profiles, especially during anaerobic growth. *S. aureus* is a facultative anaerobic bacterium and can sirvive in a low oxygen condation by fermentation or nitrate respiration and can adaptate to different levels of oxygen (Burke and Lascelles, 1975). The key regulatory system that responsible for anaerobic gene regulation is the two-component Staphylococcal respiratory response AB system (Pragman, Yarwood *et al*. 2004). Also, *S. aureus* has another system of two-component system that regulates nitrogen metabolism NreABC it has a role in anaerobic gene regulation and virulence factor expression (Schlag *et al.*, 2008). While, redox-dependent transcription repressor (Rex) also playing a crucial role in the regulation of anaerobic metabolism in *S. aureus* which employing the mechanism of sensing redox status through responding to NADH/NAD+ ratio (Pagels *et al.*, 2010). In this study the *arcR* gene encoding ArcR with the arginine deiminase pathway, genes *arcABDC* that control the use of arginine as a source of energy for growth under anaerobic conditions was significantly down-regulated after 16 h of infection. Exposure of *C. elegans* to MRSA252 has shown to down-regulationsome genes such as the *acpD* gene that involved in fatty acid and phospholipid metabolism, also putative pyruvate formate-lyase activating enzyme taht been reported to up-regulated under anaerobic conditions this enzyme is essential for energy supply when pyruvate is available (Fuchs *et al.*, 2007).

Table 21. Major groups of up-regulated genes in *S. aureus* at 16 h of *C. elegans* infection model and their predicted function

Number of genes	Predicted function
16	Capsular polysaccharide synthesis enzyme
15	Putative membrane protein
36	Synthesis enzyme protein
20	Hypothetical protein

Table 22. Top 100 up-regulated genes in *S. aureus* infection of *C. elegans* at 16 h comparing to 16 h control sample

ORF	Annotation	FC T16.vs.C16	FDR T16.vs.C16
SAR2762	ATP phosphoribosyl transferase regulatory	7.34	3.27E-14
SAR0159	Capsular polysaccharide synthesis enzyme	6.01	4.30E-34
SAR0160	Capsular polysaccharide synthesis enzyme	5.76	7.97E-19
SAR0742	Putative membrane protein	5.68	4.67E-05
SAR0415	Hypothetical protein	5.67	2.82E-04
SAR0316	Putative membrane protein	5.37	1.54E-09
SAR0158	Capsular polysaccharide synthesis enzyme	5.23	9.03E-29
SAR2449	Hypothetical protein	5.19	1.91E-18
SAR0155	Capsular polysaccharide synthesis enzyme	5.05	4.81E-21
SAR0157	Capsular polysaccharide synthesis enzyme	5.01	2.13E-24
SAR0156	Capsular polysaccharide synthesis enzyme	5.00	1.50E-23
SAR0031b	Hypothetical protein	4.98	4.76E-04
SAR0161	Capsular polysaccharide synthesis enzyme	4.90	1.29E-17
SAR0154	Capsular polysaccharide synthesis enzyme	4.87	2.12E-16
SAR2144	2-isopropylmalate synthase	4.58	1.87E-12
SAR0162	Capsular polysaccharide synthesis enzyme	4.53	5.60E-21
SAR2143	Ketol-acid reductoisomerase	4.45	1.13E-08
SAR2147	3-isopropylmalate dehydratase subunit	4.32	1.43E-10
SAR0458	Sodium neurotransmitter symporter protein	4.32	4.60E-11
SAR1000	Hypothetical protein	4.19	8.58E-03
SAR0250	Hypothetical protein	4.13	1.36E-02
SAR2648	Secretory antigen precursor	4.11	2.53E-19
SAR1405	Aspartokinase I	4.11	2.69E-17
SAR0153	Capsular polysaccharide synthesis enzyme	4.10	1.45E-14
SAR1383	Indole-3-glycerol phosphate synthase	3.95	5.93E-03
SAR0165	Capsular polysaccharide synthesis enzyme	3.94	1.13E-14
SAR0151	Capsular polysaccharide synthesis enzyme	3.92	8.22E-16
SAR1273a	Hypothetical protein	3.91	8.57E-03
SAR0152	Capsular polysaccharide synthesis enzyme	3.89	3.96E-17
SAR2450	Putative membrane protein	3.85	6.00E-16
SAR0241	Putative PTS transport system	3.80	1.04E-02
SAR2141	Acetolactate synthase large subunit	3.76	1.98E-07

Table 22-continued

ORF	Annotation	FC T16.vs.C16	FDR T16.vs.C16
SAR0996	Conserved hypothetical protein	3.74	8.19E-14
SAR2145	3-isopropylmalate dehydrogenase	3.70	6.69E-08
SAR0164	Capsular polysaccharide synthesis enzyme	3.69	8.62E-14
SAR2140	Putative dihydroxy-acid dehydratase	3.67	2.03E-05
SAR1325	Hypothetical protein	3.64	2.14E-02
SAR1459a	Hypothetical protein	3.52	3.31E-02
SAR0166	Capsular polysaccharide synthesis enzyme	3.49	5.24E-12
SAR1407	Dihydrodipicolinate synthase	3.48	1.55E-11
SAR0639	Putative halo-acid dehydrogenase-Hydrolase	3.43	4.76E-13
SAR2146	Isopropyl malate dehydratase large subunit	3.42	1.98E-08
SAR2599	Putative dioxygenase	3.42	1.57E-07
SAR0163	Capsular polysaccharide synthesis enzyme	3.38	4.95E-11
SAR1408	Dihydrodipicolinate reductase	3.18	1.10E-11
SAR0030	Hypothetical protein	3.12	4.73E-02
SAR0244	Hypothetical protein	3.07	5.33E-02
SAR1620	Putative membrane protein	3.04	4.56E-02
SAR0356	Cys/Met metabolism PLP-dependent enzyme	3.04	2.39E-03
SAR0117	Putative siderophore transport permease	2.98	1.87E-03
SAR2761	Putative ATP phosphoribosyltransferase	2.94	2.22E-03
SAR2148	Threonine dehydratase biosynthetic	2.93	9.00E-05
SAR1406	Aspartate semialdehyde dehydrogenase	2.89	6.73E-11
SAR0455	Putative membrane protein	2.88	1.29E-04
SAR1381	Anthranilate synthase component I	2.87	1.70E-02
SAR0171	Hypothetical protein	2.83	4.32E-05
SAR2752	conserved hypothetical protein	2.83	6.13E-02
SAR0731	Hypothetical protein	2.81	2.50E-12
SAR2642	Squalene synthase	2.79	1.94E-12
SAR0212	Putative membrane protein	2.74	9.53E-04
SAR1384	N-(5'phosphoribosyl) anthranilate (PRA)	2.71	9.82E-04
SAR2738	Conserved hypothetical protein	2.70	2.05E-02
SAR0495	Conserved hypothetical protein	2.70	1.09E-09
SAR0471	Glutamate synthase, large subunit	2.67	5.65E-06
SAR1411	Putative alanine racemase	2.67	7.07E-08
SAR2681	Amino acid permease family protein	2.66	4.78E-08
SAR2598	Putative phospholipase/carboxylesterase	2.61	6.04E-05
SAR2512	Putative membrane protein	2.61	4.36E-02
SAR2645	Putative glycosyl transferase	2.60	1.32E-11
SAR0238	Putative PTS multi-domain regulator	2.60	8.74E-03
SAR2682	Putative aminotransferase	2.58	1.55E-06
SAR0042	Hypothetical protein	2.53	1.13E-01

Table 22-continued

ORF	Annotation	FC T16.vs.C16	FDR T16.vs.C16
SAR0638	Putative membrane protein	2.52	6.38E-07
SAR1800	Putative soluble hydrogenase subunit	2.52	1.39E-06
SAR1108	Sortase	2.51	1.08E-01
SAR2551	Oligopeptide transporter putative ATPase	2.50	3.22E-03
SAR2651	Putative membrane protein	2.46	5.51E-04
SAR1677	Putative membrane protein	2.46	7.66E-02
SAR1380	Anthranilate synthase component	2.43	1.74E-02
SAR2507	Cation efflux family protein	2.42	1.11E-08
SAR0745	Putative membrane protein	2.41	7.32E-05
SAR1409	Putative tetrahydrodipicolinate acetyltransferase	2.39	4.23E-06
SAR2562	Putative membrane protein	2.39	2.04E-02
SAR2384	Hypothetical protein	2.38	3.56E-06

Table 23. Major groups of down-regulated genes in *S. aureus* at 16 h of *C. elegans* infection compared to non-infected control sample and their predicted function

Number of genes	Predicted function
34	Different activating enzymes
8	Membrane proteins
32	Transport ATP-binding proteins
25	Hypothetical proteins

Table 24. Top 100 down-regulated genes in *S. aureus* infection of *C. elegans* at 16 h comparing to to non-infected control sample

ORF	Annotation	FC T16.vs.C16	FDR T16.vs.C16
SAR2621	Putative membrane protein	-7.54	6.58E-17
SAR2714	Arginine deiminase	-7.49	2.34E-47
SAR2713	Putative ornithine carbamoyl-transferase	-6.74	4.94E-51
SAR1450	Putative threonine dehydratase	-6.46	6.37E-08
SAR0203	Putative phosphodiesterase	-6.28	9.82E-13
SAR2712	Arginine/ornithine antiporter	-6.18	3.58E-31
SAR2605	D-specific D-2-hydroxyacid dehydrogenase	-6.17	1.23E-23
SAR0218	Putative pyruvate formate-lyase activating enzyme	-6.15	7.44E-23
SAR0217	Formate acetyltransferase	-6.14	9.48E-15
SAR1007	Hypothetical protein	-6.09	1.16E-04
SAR2455	Putative L-lactate permease	-6.01	1.64E-22
SAR2628	Putative ATP-dependent protease ATP-binding subunit Clp	-5.85	3.26E-18
SAR2066	Hypothetical phage protein	-5.77	4.02E-04
SAR1142	Ornithine carbamoyl-transferase	-5.73	2.14E-17
SAR1832	Hypothetical protein	-5.52	3.77E-04
SAR0613	Alcohol dehydrogenase	-5.48	4.25E-16
SAR2372	Urease gamma subunit	-5.46	1.71E-03
SAR0309	Putative membrane protein	-5.08	2.54E-06
SAR0234	L-lactate dehydrogenase	-5.03	1.88E-11
SAR2711	Carbamate kinase	-5.00	8.59E-26
SAR1143	Putative carbamate kinase	-4.97	1.92E-16
SAR2569	Hypothetical protein	-4.96	9.91E-19
SAR0256	Cell wall metabolism protein	-4.95	5.85E-26
SAR2068	Putative exported protein	-4.80	2.77E-07
SAR2691	Putative betaine aldehyde dehydrogenase	-4.76	2.41E-16
SAR2071	Hypothetical phage protein	-4.75	1.14E-02
SAR0584	Hypothetical protein	-4.72	9.37E-13
SAR1451	Alanine dehydrogenase	-4.72	1.91E-05
SAR2285	Galactose-6-phosphate isomerase LacB	-4.71	1.88E-20
SAR0317	Lipase precursor	-4.70	3.50E-08
SAR2280	6-phospho-beta-galactosidas	-4.53	1.20E-16
SAR0219	Hypothetical protein	-4.51	1.12E-02

Table 24-continued

ORF	Annotation	FC T16.vs.C16	FDR T16.vs.C16
SAR2286	Galactose-6-phosphate isomerase LacA	-4.49	1.76E-15
SAR2710	Putative regulatory protein	-4.48	3.90E-19
SAR1302	Conserved hypothetical protein	-4.38	1.37E-02
SAR2284	Tagatose-6-phosphate kinas	-4.37	2.70E-16
SAR2488	Assimilatory nitrite reductase subunit	-4.36	1.52E-05
SAR1909	Hypothetical protein	-4.25	1.57E-02
SAR1136	Alpha-hemolysin precursor	-4.25	1.17E-09
SAR2296	Conserved hypothetical protein	-4.24	9.95E-11
SAR2122	Delta-hemolysin precursor	-4.17	4.15E-04
SAR2283	Tagatose 1,6-diphosphate aldolase	-4.17	2.45E-16
SAR2282	PTS system, lactose-specific IIA component	-4.14	2.68E-09
SAR0070	Potassium-transporting ATPase A	-4.13	3.06E-04
SAR2536	Glycine betaine/carnitine/choline-binding lipoprotein precursor	-3.67	9.48E-15
SAR2523	Putative membrane protein	-3.61	1.28E-12
SAR1320	Hypothetical protein	-3.61	8.38E-02
SAR0340	Putative lipoprotein	-3.57	5.40E-06
SAR1987	Putative membrane protein	-3.55	4.90E-14
SAR2695	Anaerobic ribonucleoside-triphosphate reductase	-3.55	1.07E-08
SAR2067	Hypothetical phage protein	-3.53	1.26E-13
SAR2281	PTS system, lactose-specific IIBC	-3.51	7.30E-11
SAR1932	Putative peptidyl-prolyl cis-isomerase	-3.48	1.10E-14
SAR1150	Antibacterial protein	-3.47	1.58E-02
SAR0182	Putative membrane protein	-3.47	2.88E-07
SAR0308	PfkB family carbohydrate kinas	-3.46	4.18E-06
SAR2539	Conserved hypothetical protein	-3.45	7.91E-21
SAR0528	Putative stress response Clp ATPase	-3.42	1.27E-12
SAR2535	Putative glycine betaine/carnitine/choline transport system permease protein	-3.38	6.40E-11
SAR2099	DNA-binding protein	-3.37	7.59E-09
SAR0072	Potassium-transporting ATPase C chain	-3.36	1.83E-04
SAR2072	Putative dUTP pyro-phosphatase	-3.36	1.12E-07
SAR2095	Hypothetical phage protein	-3.35	3.33E-11
SAR2519a	Hypothetical protein	-3.35	1.45E-04
SAR1974	Putative response regulator	-3.33	1.38E-17
SAR0232	Putative membrane protein	-3.33	3.55E-03
SAR0261	Putative nitric oxide reductase	-3.33	4.26E-07
SAR0419	Hypothetical protein	-3.33	1.01E-01
SAR1658	GrpE protein- Hsp-70 cofactor	-3.33	2.16E-10
SAR1657	Chaperone protein	-3.32	1.88E-09
SAR0369	Conserved hypothetical protein	-3.27	1.51E-01
SAR2715	Arginine repressor family protein	-3.25	1.63E-06
SAR2694	Putative anaerobic ribonucleotide reductase activating protein	-3.24	6.28E-06
SAR2091	Putative exported protein	-3.24	1.07E-01

Table 24-continued

ORF	Annotation	FC T16.vs.C16	FDR T16.vs.C16
SAR1322	Hypothetical protein	-3.24	1.00E-01
SAR2094	Hypothetical phage protein	-3.23	1.50E-05
SAR0233	Flavohemo protein	-3.22	7.28E-13
SAR1709	Hypothetical protein	-3.22	1.05E-12
SAR2117	10 kDa chaperonin	-3.21	1.38E-15
SAR2096	Putative anti repressor	-3.21	1.61E-19
SAR2089	Hypothetical phage protein	-3.20	7.98E-09
SAR1659	Heat-inducible transcription repressor	-3.20	2.95E-08
SAR0130	NAD dependent epimerase/dehydratase family protein	-3.18	2.39E-02
SAR0525	Putative DNA-binding protein	-3.16	2.19E-17
SAR2074	Hypothetical phage protein	-3.16	1.42E-01
SAR0526	Conserved hypothetical protein	-3.16	4.15E-13
SAR1975	Histidine kinase sensor	-3.16	9.92E-16
SAR2538	Glycine betaine/carnitine/choline Transport ATP-binding protein	-3.15	2.50E-15

Table 25. Major groups of up-regulated genes of *S. aureus* at 40 h of *C. elegans* infection comparing to non-infected control sample and their predicted function

Number of genes	Predicted function
19	Hypothetical protein
16	Capsular polysaccharide synthesis enzyme
11	Putative membrane protein
3	Membrane lipoprotein

Table 26. List of the most up-regulated genes of *S. aureus* at 40 h after infection of *C. elegans* comparing to 16 h control sample

ORF	Annotation	FC T40.vs.C16	FDR T40.vs.C16
SAR2762	ATP phosphoribosyltransferase regulatory	8.83	9.62E-20
SAR0159	Capsular polysaccharide synthesis enzyme	6.59	2.76E-38
SAR0154	Capsular polysaccharide synthesis enzyme	6.50	3.34E-24
SAR0984	Transposase	6.38	1.14E-04
SAR0160	Capsular polysaccharide synthesis enzyme	6.33	6.02E-21
SAR0155	Capsular polysaccharide synthesis enzyme	6.31	1.52E-28
SAR0158	Capsular polysaccharide synthesis enzyme	6.14	7.29E-36
SAR0156	Capsular polysaccharide synthesis enzyme	6.10	5.88E-31
SAR0157	Capsular polysaccharide synthesis enzyme	6.05	1.23E-31
SAR0742	Putative membrane protein	5.84	1.79E-04
SAR0153	Capsular polysaccharide synthesis enzyme	5.83	6.76E-24
SAR2648	Secretory antigen precursor	5.81	2.07E-32
SAR0152	Capsular polysaccharide synthesis enzyme	5.62	4.09E-29
SAR0151	Capsular polysaccharide synthesis enzyme	5.44	2.57E-25
SAR0451	Hypothetical protein	4.92	1.87E-02
SAR2761	Putative ATP phosphoribosyltransferase	4.90	1.15E-06
SAR0161	Capsular polysaccharide synthesis enzyme	4.83	6.97E-17
SAR1273a	Hypothetical protein	4.81	5.69E-03
SAR0393	Hypothetical protein	4.78	9.59E-03
SAR0162	Capsular polysaccharide synthesis enzyme	4.77	2.64E-22
SAR0471	Glutamate synthase, large subunit	4.50	2.74E-12
SAR0011	Putative membrane protein	4.39	1.30E-02
SAR0165	Capsular polysaccharide synthesis enzyme	4.37	1.00E-16
SAR1380	Anthranilate synthase component	4.35	9.68E-05
SAR1381	Anthranilate synthase component I	4.26	8.16E-04
SAR1383	Indole-3-glycerol phosphate synthase	4.25	9.83E-03
SAR0166	Capsular polysaccharide synthesis enzyme	4.24	1.45E-15
SAR2683	Putative membrane protein	4.09	6.63E-21
SAR1325	Hypothetical protein	3.94	4.00E-02
SAR0706	Putative membrane lipoprotein	3.94	4.25E-02
SAR0458	Sodium: neurotransmitter symporter	3.92	2.85E-09
SAR0164	Capsular polysaccharide synthesis enzyme	3.90	2.53E-14

Table 26-continued

ORF	Annotation	FC T40.vs.C16	FDR T40.vs.C16
SAR0415	Hypothetical protein	3.88	3.73E-02
SAR0055	DNA repair protein RadC	3.87	4.48E-02
SAR1866	CrcB-like protein	3.87	4.48E-02
SAR2517	Putative dethiobiotin synthetase	3.87	4.48E-02
SAR0472	Glutamate synthase, small subunit	3.80	3.91E-09
SAR2745	Putative capsule synthesis protein	3.70	4.60E-02
SAR0731	Hypothetical protein	3.66	2.89E-18
SAR2643	Squalene desaturase	3.65	1.44E-16
SAR1040	Putative phosphor ribosylaminoimidazole succino-carboxamide synthase	3.62	6.05E-08
SAR2384	Hypothetical protein	3.54	6.55E-11
SAR1401	ABC transporter permease protein	3.54	2.23E-08
SAR1283	Putative membrane protein	3.94	3.44E-14
SAR2645	Putative glycosyl transferase	3.04	2.75E-14
SAR0675	Putative exported protein	3.03	2.11E-14
SAR0010	Putative membrane protein	3.03	4.06E-02
SAR2758	Putative imidazole glycerol-phosphate dehydratase	3.02	1.87E-03
SAR1284	Glutamine synthetase	3.01	6.26E-09
SAR1041	Conserved hypothetical protein	3.00	1.25E-07
SAR1337	Putative aspartate kinas	2.96	1.36E-05
SAR1855	Arsenical resistance operon repressor	2.94	1.22E-01
SAR0316	Putative membrane protein	2.93	1.30E-03
SAR1922	Hypothetical protein	2.91	2.68E-02
SAR0171	Hypothetical protein	2.88	3.22E-04
SAR2145	3-isopropylmalate dehydrogenase	2.86	4.55E-05
SAR1382	Anthranilate phosphoribosyltransferase	2.85	1.11E-04
SAR0241	Putative PTS transport system	2.85	1.54E-01
SAR2143	Ketol-acid reductoisomerase	2.82	2.40E-04
SAR2647	Putative membrane protein	2.82	5.97E-10
SAR0200	Putative transport system permease	2.80	2.42E-02
SAR1910	Putative membrane protein	2.79	1.76E-01
SAR0356	Cys/Met metabolism PLP enzyme	2.78	1.44E-02
SAR0117	Putative siderophore transport permease	2.78	1.58E-02
SAR0066	Putative transposase	2.77	1.38E-01
SAR1285	Hypothetical protein	2.75	1.79E-06
SAR0589	Putative amino acid permease	2.74	6.54E-13
SAR2144	2-isopropylmalate synthase	2.74	3.46E-05
SAR1402	Phosphate-binding lipoprotein	2.71	6.47E-06
SAR2147	3-isopropylmalate dehydratase subunit	2.71	2.89E-03
SAR1400	ABC transporter permease protein	2.69	1.70E-05
SAR2750	Intercellular adhesion protein	2.68	5.70E-02
SAR2059	Hypothetical phage protein	2.67	9.91E-02
SAR0854	Hypothetical protein	2.67	4.69E-04
SAR0142	Binding-protein-dependent transport systems membrane component	2.66	6.44E-03
SAR1800	Putative soluble hydrogenase subunit	2.65	1.90E-06

Table 26-continued

ORF	Annotation	FC T40.vs.C16	FDR T40.vs.C16
SAR1930a	Hypothetical protein	2.62	1.06E-01
SAR0128	Putative membrane protein	2.59	1.22E-03
SAR2457	Putative lipoprotein	2.58	1.60E-10
SAR1921	Enterotoxin	2.57	6.79E-02
SAR1000	Hypothetical protein	2.56	2.02E-01
SAR2781	ABC transporter ATP-binding protein	2.55	3.64E-02
SAR0282	Conserved hypothetical protein	2.54	2.18E-01
SAR1598	Arginine repressor	2.53	4.36E-11
SAR1943	Putative membrane protein	2.51	3.10E-05
SAR0036	Putative membrane protein	2.47	1.34E-01
SAR2646	Putative phytoene dehydrogenase related protein	2.47	2.57E-08
SAR2681	Amino acid permease family protein	2.42	1.81E-06

Table 27. Major groups of down-regulated genes in *S. aureus* at 40 h of *C. elegans* infection comparing to 16 h control sample model and their predicted function

Number of genes	Predicted function
35	Hypothetical protein
10	ATP-binding transporter protein
10	Amino acids metabolic enzymes
8	Putative membrane protein

Table 28. The most down-regulated genes in *S. aureus* at 40h of *C. elegans* infection comparing to 16 h control sample

ORF	Annotation	FC T40.vs.C16	FDR T40.vs.C16
SAR2621	Putative membrane protein	-9.80	3.70E-10
SAR2132	Putative membrane protein	-8.88	8.65E-09
SAR2715	Arginine repressor family protein	-8.83	3.28E-07
SAR2070	Hypothetical phage protein	-8.72	4.47E-09
SAR1451	Alanine dehydrogenase	-8.26	1.62E-04
SAR2055	Hypothetical phage protein	-7.53	7.38E-05
SAR2793	Putative membrane protein	-7.52	1.77E-04
SAR2054	Hypothetical phage protein	-7.51	4.61E-04
SAR2605	D-specific D-2-hydroxyacid dehydrogenase	-7.34	9.41E-25
SAR2052	Hypothetical phage protein	-7.31	1.98E-04
SAR0173	Putative ABC transporter ATP-binding protein	-7.22	2.95E-04
SAR2061	Hypothetical phage protein	-7.20	1.85E-04
SAR0989	Hypothetical protein	-7.17	4.01E-04
SAR2068	Putative exported protein	-7.10	3.00E-04
SAR2040	Autolysis	-7.01	4.12E-04
SAR2714	Arginine deiminase	-6.74	2.33E-33
SAR2373	Urease beta subunit	-6.69	4.07E-03
SAR2088	Hypothetical phage protein	-6.60	2.75E-03
SAR2488	Assimilatory nitrite reductase small subunit	-6.59	2.16E-03
SAR2053	Hypothetical phage protein	-6.42	6.21E-03
SAR0907	Hypothetical protein	-6.31	1.45E-02
SAR0423	Exotoxin	-6.25	1.11E-02
SAR0047	Conserved hypothetical protein	-6.22	5.85E-03
SAR0787	FecCD transport family protein	-6.20	9.82E-03
SAR2076	Hypothetical phage protein	-6.19	1.07E-02
SAR1007	Hypothetical protein	-6.09	1.43E-02
SAR2163	Putative potassium-transporting ATPase	-6.06	1.04E-02
SAR0174	Putative lipoprotein	-5.98	1.27E-02
SAR2712	Arginine/ornithine antiporter	-5.94	7.47E-23
SAR2711	Carbamate kinas	-5.94	3.06E-27
SAR0788	FecCD transport family protein	-5.92	1.97E-02
SAR1141	Exotoxin	-5.89	2.44E-02
SAR0218	Putative pyruvate formate-lyase enzyme	-5.89	2.12E-19

Table 28-continued

ORF	Annotation	FC T40.vs.C16	FDR T40.vs.C16
SAR2066	Hypothetical phage protein	-5.77	2.59E-02
SAR2631	Hypothetical protein	-5.74	3.97E-02
SAR0294	Conserved hypothetical protein	-5.72	3.08E-02
SAR2046	Hypothetical phage protein	-5.69	3.40E-02
SAR0428	Exotoxin	-5.69	2.21E-02
SAR2628	Putative ATP-dependent protease ATP-binding subunit Clp	-5.66	2.51E-16
SAR0435	Exotoxin	-5.52	2.55E-02
SAR1832	Hypothetical protein	-5.52	3.04E-02
SAR2058	Hypothetical phage protein	-5.50	2.70E-02
SAR2372	Urease gamma subunit	-5.46	4.75E-02
SAR0256	Cell wall metabolism protein	-5.45	4.88E-22
SAR2710	Putative regulatory protein	-4.88	4.73E-18
SAR1079	Putative manganese transport protein	-4.86	2.54E-23
SAR2569	Hypothetical protein	-4.79	2.55E-15
SAR0439	Putative lipoprotein	-4.76	1.28E-01
SAR1450	Putative threonine dehydratase	-4.75	3.50E-04
SAR2071	Hypothetical phage protein	-4.75	1.14E-01
SAR2731	Conserved hypothetical protein	-4.71	9.30E-06
SAR0613	Alcohol dehydrogenase	-4.71	2.51E-11
SAR2117	10 kDa chaperonin	-4.64	1.30E-21
SAR1907	Serine protease	-4.63	1.08E-01
SAR2742	Acetyltransferase (GNAT) family protein	-4.63	1.16E-01
SAR0416	Putative transposes	-4.63	1.05E-01
SAR2691	Putative betaine aldehyde dehydrogenase	-4.63	6.09E-14
SAR0122	Putative transport protein	-4.63	1.21E-01
SAR1657	Chaperone protein	-4.62	7.12E-15
SAR2743	Putative capsule synthesis protein	-4.59	1.23E-01
SAR1686	Putative biotin carboxyl carrier protein of acetyl-CoA carboxylase	-4.58	1.30E-01
SAR0641	ABC transporter extracellular binding	-4.58	2.84E-14
SAR2044	Hypothetical protein	-4.57	1.23E-01
SAR1350	Putative membrane protein	-4.57	7.37E-05
SAR0361	Hypothetical protein	-4.56	4.14E-07
SAR0120	Putative ornithine cyclodeaminase	-4.55	1.16E-01
SAR1449	Amino acid permease	-4.55	2.30E-04
SAR2768	Conserved hypothetical protein	-4.54	1.37E-01
SAR1143	Putative carbamate kinas	-4.53	5.18E-10
SAR2085	Hypothetical phage protein	-4.49	4.72E-18
SAR0119	Pyridoxal-phosphate dependent enzyme	-4.49	1.33E-01
SAR1625	Conserved hypothetical protein	-4.46	1.59E-01
SAR0941	Hypothetical protein	-4.46	1.21E-01
SAR1658	GrpE protein (Hsp-70 cofactor)	-4.45	7.64E-15
SAR0111	Putative myosin-crossreactive antigen	-4.44	7.44E-12
SAR2051	Hypothetical phage protein	-4.44	1.53E-01
SAR1330	Putative membrane protein	-4.43	1.31E-01
SAR0286	Hypothetical protein	-4.41	1.51E-01

Table 28-continued

ORF	Annotation	FC T40.vs.C16	FDR T40.vs.C16
SAR2072	Putative dUTP pyrophosphatase	-4.41	1.31E-01
SAR2694	Putative anaerobic ribonucleotide reductase Activating protein	-4.37	1.51E-01
SAR2658	TetR family regulatory protein	-4.36	1.79E-05
SAR0635	Putative membrane protein	-4.36	4.23E-07
SAR0114	Immunoglobulin G binding protein A precursor	-4.34	9.50E-05
SAR1109	Conserved hypothetical protein	-4.33	1.42E-01
SAR2087	Hypothetical phage protein	-4.31	6.47E-08
SAR1909	Hypothetical protein	-4.25	1.45E-01
SAR0642	ABC transporter permease protein	-4.25	4.72E-05
SAR0209	Putative oxidoreductase	-4.23	1.53E-01
SAR2537	Putative glycine betaine/carnitine/choline Transport system permease protein	-4.21	4.34E-13

Table 29. Major groups of up-regulated genes in *S. aureus* at 40 h of *C. elegans* infection compared to 16 h of infection sample and their predicted function

Number of genes	Predicted function
25	Metabolic enzyme
10	Transport system proteins
9	Putative membrane protein
4	Capsule synthesis protein

Table 30. Top 100 up-regulated genes in *S. aureus* infection of *C. elegans* at 40 h comparing to 16 h of infection sample

ORF	Annotation	FC T40.vs.T16	FDR T40.vs.T16
SAR0451	Hypothetical protein	6.79	7.54E-03
SAR1402	Phosphate-binding lipoprotein	4.85	1.35E-12
SAR1302	Conserved hypothetical protein	4.83	2.21E-02
SAR0219	Hypothetical protein	4.83	2.23E-02
SAR0393	Hypothetical protein	4.78	4.25E-02
SAR1126	Hypothetical protein	4.75	5.78E-02
SAR1401	ABC transporter permease protein	4.71	3.06E-10
SAR1399	ABC transporter ATP-binding protein	4.40	3.09E-11
SAR2354	Putative molybdopterin-synthase	4.32	8.65E-03
SAR0984	Transposase	4.22	1.36E-02
SAR0203	Putative phosphodiesterase	4.15	8.55E-04
SAR1400	ABC transporter permease protein	3.95	3.90E-08
SAR1320	Hypothetical protein	3.94	1.10E-01
SAR0706	Putative membrane protein	3.94	1.12E-01
SAR2519	ABC transporter ATP-binding protein	3.90	3.43E-04
SAR0419	Hypothetical protein	3.87	1.10E-01
SAR2613	Putative membrane protein	3.87	4.08E-05
SAR0062	Hypothetical protein	3.87	1.14E-01
SAR2243	Hypothetical protein	3.87	1.14E-01
SAR0282	Conserved hypothetical protein	3.87	1.16E-01
SAR0055	DNA repair protein RadC	3.87	1.18E-01
SAR1866	CrcB-like protein	3.87	1.18E-01
SAR2517	Putative dethiobiotin synthetase	3.87	1.18E-01
SAR1398	Putative phosphate transport protein	3.85	4.22E-09
SAR0317	Lipase precursor	3.84	3.30E-06
SAR0309	Putative membrane protein	3.60	1.19E-02
SAR0070	Potassium-transporting ATPase A chai	3.47	1.69E-02
SAR1896	Hypothetical protein	3.42	6.85E-03
SAR0766	Glutamine amidotransferase protein	3.29	6.93E-02
SAR0105	Phosphatidylinositol phosphodiesterase	3.01	6.24E-02
SAR0142	Binding-protein-dependent transport systems membrane component	2.98	7.18E-03
SAR1038	Putative phosphoribosylaminoimidazole carboxylase catalytic subunit	2.95	4.69E-04

Table 30-continued

ORF	Annotation	FC T40.vs.T16	FDR T40.vs.T16
SAR0011	Putative membrane protein	2.91	8.94E-02
SAR0431	Exotoxin	2.83	1.44E-03
SAR0308	PfkB family carbohydrate kinas	2.81	5.38E-03
SAR0228	Putative glutamine amidotransferase	2.68	1.22E-03
SAR2745	Putative capsule synthesis protein	2.65	1.92E-01
SAR2349	MarR family regulatory protein	2.63	6.60E-03
SAR0580	Putative AMP-binding enzyme	2.61	2.93E-02
SAR1012	Hypothetical protein	2.60	1.27E-01
SAR1930a	Hypothetical protein	2.58	1.59E-01
SAR1132	Hypothetical protein	2.56	7.47E-02
SAR0437	Putative exported protein	2.54	4.31E-05
SAR1041	Conserved hypothetical protein	2.45	3.64E-05
SAR1294	Putative membrane protein	2.25	8.40E-02
SAR0327	Putative PTS transport system protein	2.21	1.05E-01
SAR1855	Arsenical resistance operon repressor	2.18	2.81E-01
SAR2477	Conserved hypothetical protein	2.17	8.27E-02
SAR2285	Galactose-6-phosphate isomerase LacB	2.14	1.00E-03
SAR2758	Putative imidazole glycerol-phosphate dehydratase	2.11	3.36E-02
SAR0411	Hypothetical protein	2.11	4.16E-04
SAR0617	Putative DNA repair protein	2.10	5.02E-02
SAR1042	Putative phosphoribosyl formyl Glycinamidine synthase	2.08	5.18E-03
SAR0335	Putative luciferase-like monooxygenase	2.06	1.20E-06
SAR1289	Putative exported protein	2.05	7.89E-03
SAR0336	NADH-dependent FMN reductase	1.99	5.16E-04
SAR0626	Putative membrane protein	1.99	7.46E-02
SAR0934	Halo-acid dehalogenase-like hydrolase	1.96	2.02E-07
SAR2761	ATP phosphoribosyltransferase	1.96	3.64E-02
SAR1380	Anthranilate synthase component	1.93	6.69E-02
SAR1225	SMF family protein	1.91	3.76E-03
SAR2162	Conserved hypothetical protein	1.91	9.05E-03
SAR2792	Putative membrane protein	1.91	6.13E-02
SAR2289a	Hypothetical protein	1.91	2.11E-01
SAR0146	Putative exported protein	1.91	2.35E-02
SAR2744	Putative capsule synthesis protein	1.90	1.85E-01
SAR0846	Putative exported protein	1.89	1.50E-01
SAR0185	Putative N-acetyl-gamma-glutamyl-phosphate reductase	1.88	1.23E-01
SAR1772	Alkaline phosphatase synthesis Transcriptional regulatory protein	1.88	1.21E-06
SAR1142	Ornithine carbamoyl transferase	1.87	2.87E-02
SAR0967	Hypothetical protein	1.87	3.07E-01
SAR2509	Gamma-hemolysin component A precursor	1.84	7.69E-03

Table 30-continued

ORF	Annotation	FC T40.vs.T16	FDR T40.vs.T16
SAR0471	Glutamate synthase, large subunit	1.83	9.57E-02
SAR0200	Putative transport system permease	1.83	9.75E-02
SAR2251	Putative transposase	1.81	1.68E-01
SAR2122	Delta-hemolysin precursor	1.79	6.32E-06
SAR2478	MerR family regulatory protein	1.78	7.12E-03
SAR2624	Putative exported protein	1.76	2.80E-01
SAR0961	Putative transposase	1.75	8.16E-03
SAR2297b	Hypothetical protein	1.74	4.40E-01
SAR1910	Putative membrane protein	1.74	3.50E-04
SAR0153	Capsular polysaccharide synthesis enzyme	1.73	4.63E-05
SAR0152	Capsular polysaccharide synthesis enzyme	1.73	4.20E-01
SAR0131	Putative sugar transferase	1.73	3.88E-01
SAR1450	Putative threonine dehydratase	1.71	3.76E-05
SAR2648	Secretory antigen precursor	1.70	3.56E-02

Table 31. Major groups of down-regulated genes in *S. aureus* at 40 h of *C. elegans* infection comparing to 16 h of infection sample and their predicted function

Number of genes	Predicted function
46	Hypothetical proteins
14	Different enzymes
15	Putative membrane proteins
13	Transport family proteins

Table 32. Top 100 down-regulated genes in *S. aureus* infection of *C. elegans* at 40 h comparing to 16 h of infection sample

ORF	Annotation	FC T40.vs.T16	FDR T40.vs.T16
SAR0788	FecCD transport family protein	-6.91	3.41E-03
SAR0047	Conserved hypothetical protein	-6.71	2.01E-03
SAR2061	Hypothetical phage protein	-6.70	1.46E-03
SAR0787	FecCD transport family protein	-6.68	3.97E-03
SAR0989	Hypothetical protein	-6.64	2.40E-03
SAR2132	Putative membrane protein	-6.51	2.27E-03
SAR2046	Hypothetical phage protein	-6.20	1.80E-02
SAR2054	Hypothetical phage protein	-6.08	1.47E-02
SAR2055	Hypothetical phage protein	-5.78	1.69E-02
SAR2557	Conserved hypothetical protein	-5.73	2.11E-02
SAR0173	Putative ABC transporter ATP-binding protein	-5.70	2.10E-02
SAR0122	Putative transport protein	-5.66	4.46E-02
SAR1298	Hypothetical protein	-5.65	4.98E-02
SAR2715	Arginine repressor family protein	-5.58	2.62E-02
SAR0936	Hypothetical protein	-5.43	1.20E-01
SAR0174	Putative lipoprotein	-5.35	4.26E-02
SAR0143	Binding-protein-dependent transport system membrane component	-5.33	7.66E-02
SAR1677	Putative membrane protein	-5.32	8.55E-02
SAR2163	Putative potassium-transporting ATPase C	-5.27	4.65E-02
SAR2562	Putative membrane protein	-5.23	5.83E-02
SAR2631	Hypothetical protein	-5.22	7.48E-02
SAR0250	Hypothetical protein	-5.21	1.09E-01
SAR1912	Transposase	-5.21	1.18E-01
SAR0793	Hypothetical protein	-5.20	7.52E-02
SAR2230	Putative membrane protein	-5.18	8.88E-02
SAR2391a	Putative membrane protein	-5.17	9.75E-02
SAR2058	Hypothetical phage protein	-5.14	5.39E-02
SAR0428	Exotoxin	-5.03	6.52E-02
SAR2373	Urease beta subunit	-5.02	7.15E-02
SAR0031b	Hypothetical protein	-4.98	9.53E-02
SAR2552	Oligopeptide transporter putative membrane permease	-4.96	1.21E-01

Table 32-continued

ORF	Annotation	FC T40.vs.T16	FDR T40.vs.T16
SAR2070	Hypothetical phage protein	-4.95	6.01E-02
SAR0090	Hypothetical protein	-4.92	1.09E-01
SAR0120	Putative ornithine cyclodeaminase	-4.92	8.60E-02
SAR1620	Putative membrane protein	-4.90	1.26E-01
SAR0242	Putative PTS transport system, IIC	-4.89	8.70E-02
SAR2752	Conserved hypothetical protein	-4.87	1.29E-01
SAR1625	Conserved hypothetical protein	-4.85	1.24E-01
SAR1107	Iron/heme permease	-4.75	1.25E-01
SAR1619	Putative exported protein	-4.74	1.42E-01
SAR2768	Conserved hypothetical protein	-4.72	1.23E-01
SAR2040	Autolysin	-4.71	8.74E-02
SAR0907	Hypothetical protein	-4.70	1.17E-01
SAR0057	Hypothetical protein	-4.68	1.49E-01
SAR0423	Exotoxin	-4.40	1.42E-01
SAR2450	Putative membrane protein	-4.38	3.05E-12
SAR1108	Sortase	-4.36	1.89E-01
SAR0100	Putative membrane protein	-4.33	1.78E-01
SAR1981	Hypothetical protein	-4.32	2.12E-01
SAR0941	Hypothetical protein	-4.23	1.60E-01
SAR0244	Hypothetical protein	-4.16	2.13E-01
SAR2041	Holi	-4.15	2.02E-01
SAR1907	Serine protease	-4.10	1.87E-01
SAR2060	Hypothetical phage protein	-4.07	1.92E-01
SAR0285	Hypothetical protein	-4.06	1.92E-01
SAR0435	Exotoxin	-4.01	1.84E-01
SAR2053	Hypothetical phage protein	-3.98	1.95E-01
SAR2741	Hypothetical protein	-3.91	2.08E-01
SAR0643	ABC transporter ATP-binding protein	-3.91	7.11E-13
SAR0642	ABC transporter permease protein	-3.84	5.15E-11
SAR2520	Hypothetical protein	-3.81	2.37E-01
SAR2088	Hypothetical phage protein	-3.79	2.25E-01
SAR0209	Putative oxidoreductase	-3.78	1.04E-08
SAR0133	Putative membrane protein	-3.76	2.50E-01
SAR0641	ABC transporter extracellular binding protein	-3.75	1.81E-10
SAR2112	Putative transposase	-3.68	2.53E-01
SAR1330	Putative membrane protein	-3.68	2.51E-01
SAR0208	Putative sugar transport system permease	-3.60	8.69E-05
SAR0376	Hypothetical protein	-3.59	2.71E-01
SAR0460	Putative Cys/Met metabolism PLP-dependent enzyme	-3.58	2.21E-07
SAR1451	Alanine dehydrogenase	-3.54	2.70E-01
SAR1109	Conserved hypothetical protein	-3.54	2.73E-01
SAR0627	Hypothetical protein	-3.53	3.19E-01
SAR1459a	Hypothetical protein	-3.52	3.18E-01
SAR2449	Hypothetical protein	-3.44	1.50E-06
SAR0805	Conserved hypothetical protein	-3.41	3.16E-01

Table 32-continued

ORF	Annotation	FC T40.vs.T16	FDR T40.vs.T16
SAR0286	Hypothetical protein	-3.40	3.16E-01
SAR2582	Putative gluconate permease	-3.33	2.52E-12
SAR1873	Conserved hypothetical protein	-3.31	2.93E-03
SAR0459	Pyridoxal-phosphate dependent enzyme	-3.27	1.36E-04
SAR1317	Hypothetical protein	-3.27	3.38E-01
SAR2076	Hypothetical phage protein	-3.26	3.22E-01
SAR2266	FecCD transport family protein	-3.25	4.10E-03
SAR0260	Holin-like protein	-3.22	1.88E-06
SAR2466	PTS system, sucrose-specific IIBC component	-3.21	1.94E-07
SAR1141	Exotoxin	-3.15	3.43E-01
SAR0259	Holin-like protein	-3.13	2.82E-04

5.2.6 Differential expression analysis

Approx. 50 million reads were obtained from the mixture of each RNA sample, reads were aligned to the combined genome sequences using Tophat (version 2.1.0), then gene expressions were calculated from reads alignment files. The count data were split into two sub sets: one containing transcripts belong to *S. aureus*, other containing transcripts for *C. elegans*. The two sub data sets were analysed separately. The differential genes expression analyses were applied to the two sub count data sets, respectively. The estimated Fold Changes (FC) was tested in edgeR using a Likelihood-Ratios (LR) test. P-values associated with FC were adjusted for multiple testing using the False Discovery Rate (FDR) approach (Benjamini and Hochberg, 1995). Significantly differentially expressed genes were defined as those with FDR-adjusted P-value < 5%.

5.2.6.1 Identification of differentially expressed (DE) genes of *S. aureus* MRSA252 in *C. elegans* infection model by using RNA-seq technology

The transcriptome analyses of the three conditions in this experiment showed a group of *S. aureus* MRSA252 genes that were expressed during infection of *C. elegans*, the three times shared many genes that were expressed with different levels, the samples of early and late stages of infection have shared 307 expressed genes. However, 44 genes similar in expression between the three transcriptome conditions. The early time-point of post-infection (16 h) revealed that 799 genes were expressed in total in this condition, 349 transcripts up-regulation and 450 transcripts were down-regulation. At the late stage of infection the *S. aureus* transcriptome had 782 transcripts differentially expressed with 324 transcripts up-regulation and 458 transcripts were down-regulation. Seventy-five transcripts were expressed only during 16 and 40 h. The number of DE genes for *S. aureus* are listed in table 33; also differentially expressed genes common to the different sampling times are summarized in a Venn diagram (figure 24) which produced by Venny 2.1 software from http://bioinfogp.cnb.csic.es/tools/venny/.

Table 33. Number of Differentially Expressed genes of *S. aureus* MRSA252 during the *C. elegans* infection model

Regulation	T16.vs.C16	T40.vs.C16	T40.vs.T16
Differentially Expressed	799	782	307
Up-regulated DE	349	324	118
Down- regulated DE	450	458	189

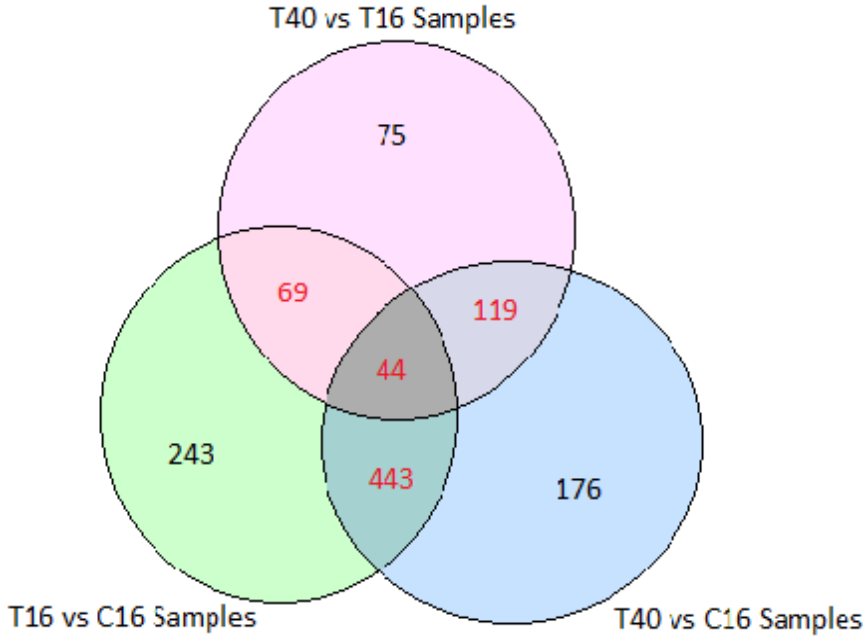

Figure 24. Proportional Venn diagram summarizing common differentially expressed genes of *S. aureus* MRSA252 during the infection model of *C. elegans*

Samples taken 16 and 40 h post-infection conditions compared with the 16 h control, the overlap area indicates the common differentially expressed genes among three conditions. Venn diagram produced by Venny 2.1 software from http://bioinfogp.cnb.csic.es/tools/venny/.

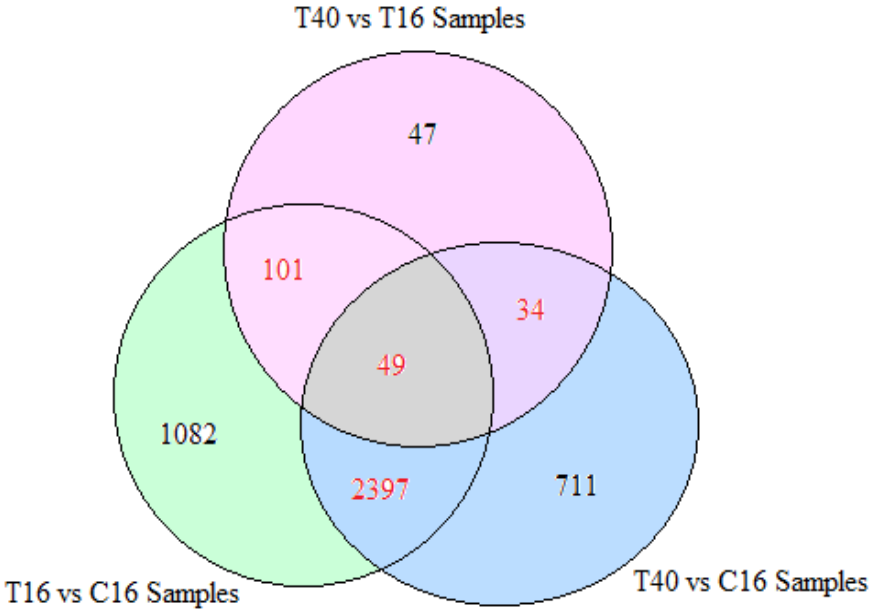

Figure 25. Proportional Venn diagram summarizing common differentially expressed genes of *C. elegans* during the infection model

Analysis of RNA-Sequencing data of 16 and 40 h post-infection conditions compared with the 16 h control; the overlap area indicates the common differentially expressed genes among three conditions.

5.2.6.2 Differentially expressed genes of *C. elegans* during the infection model

Expression levels of *C. elegans* transcripts were calculated to find out the differentially expressed genes, *C. elegans* transcriptional profiling reveals ~ 18,000 transcripts were identified. Initially, in the 16 h of infection condition compared with the non-infected nematodes 3629 differentially expressed genes were profiled in total, including 1776 transcripts that were down-regulated and 1853 transcripts that were up-regulated, while 1880 transcripts were up-regulated and 1311 transcripts down-regulated in the late stage of infection (40h). Meanwhile, 231 differentially expressed genes were expressed in both times of infection with 53 most down-regulated and 178 up-regulated transcripts. The number of *C. elegans* DE genes is listed in table 34, also the differentially expressed genes common to different times of infection are summarized in a Venn diagram (figure 25).

Table 34. Number of Differentially Expressed genes of *C. elegans*

Regulation	T16.vs.C16	T40.vs.C16	T40.vs.T16
Differentially Expressed	3629	3191	231
Up-regulated DE	1853	1880	178
Down- regulated DE	1776	1311	53

5.2.8 Analysis of *C. elegans* gene expression changes induced by *S. aureus* infection

The bacterial pathogen *S. aureus* colonize the intestine of nematodes and caused the alteration of expressed genes, over 2,000 transcripts were expressed in up and down-regulated manner, the top 100 genes up and down-regulated in the *C. elegans* transcripts were identified to investigate the transcriptome complexity changes that accompany infection and show the innate immune response of *C. elegans*. Both infection times showed detailed transcriptional response of *C. elegans* to *S. aureus* infection, the comparison of this transcriptome data characterize the most transcriptional changes in *C. elegans* genes that accompany infection, quantitative transcript levels of the top 100 down and up regulated genes are summarized in tables 35-46. The biological variation for 16 h time-point of infection were compared to 16 h of non-infected sample, 40 h of infection compared to 16 h of non-infected sample and 16 h of infection were compared to 40 h of infection sample.

Among the 100 genes most up-regulated after 16 h (tables 35 and 36) there were a group of cuticle collagen protein genes which are a common component of *C. elegans* response to the infection. Several protein-protein interaction genes are known to be involved in the enzyme activities and antibody antigen interaction were identified in significant up-regulated levels. Also 6 zygotic transcript proteins were identified as up-regulated in the early stage of infection.

In the late stage of *C. elegans* infection (40h) the cellular activities of nematodes was decreased as most of metabolic and developments process responsible genes were showed to be down-regulated compared with the early stage activities of those genes. Three cuticle collagen genes were activated during the late time of infection, also few of protein-protein interactions *fbxa* genes were upregulated after 40 h (tables 37 and 38). The transcriptional changes between the two time-points of infection revealed the unknown protein functions were the most up-regulated transcripts as 38 expressed genes of hypothetical or unknown function proteins (table 39 and 40). Thirty protein-protein interaction genes were up-regulated and 6 zygotic transcript transcripts showed up-regulated level at the late phase of bacterial infection.

The early phase of infection showed several down-regulated genes (tables 41 and 42), 30 of cuticle collagen transcripts had reduced levels of expression, 2 C-type lectin genes were

among the down-regulated transcripts and several genes of unknown function were decreased. In table 44, among the top 100 *C. elegans* down-regulated genes after 40 h of infection the transcriptome showed decline in the metabolic enzymes activities and metals ions binding proteins biosynthesis, 13 signaling receptor genes were among the most down-regulated transcripts, also 2 C-type lectin transcripts had low level of expression in late infection stage comparing with non-infected control.

The top 100 up-or-down-regulated genes of *C. elegans* after 16 and 40 h of infection with *S. aureus* MRSA252 are summarized in table 47. The majority of these induced genes were predicted as undefined function but their induction in this particular time of infection suggests an important role in the *C. elegans* innate immune system. The up-regulation levels for ~ 90% of them were during the early infection stage, while these genes had no down-regulation in the late infection stages. Fourteen embryogenesis and growth factor genes were highly induced and showed high level of up-regulation at the early infection time and down-regulation during the late stage of infection. A set of heat shock protein had a high level of transcripts during the course of *S. aureus* infection stages. Two cytochrome P450s encoding genes had a low expression level as they showed down-regulation in the early time of *C. elegans* infection.

Table 35. Major groups of up-regulated genes in *C. elegans* at 16 h of infection model compared with the non-infected 16 h control sample their predicted function

Number of genes	Predicted function
36	Protein-protein interactions
23	Hypothetical proteins
6	Zygotic transcripts
5	Unknown
3	Zinc ion binding activity

Table 36. Top 100 *C. elegans* genes up-regulated after 16 h of infection compared with the non-infected 16 h control sample

Name	Function	FC T16.vs.C16	FDR T16.vs.C16
sdz-4	Zygotic transcript	9.25	1.53E-11
hnd-1	Transcription factor	9.10	1.66E-12
sdz-5	Zygotic transcript	8.98	1.00E-10
T04C12.1	Hypothetical protein	8.93	2.18E-13
ZK899.6	Hypothetical protein	8.86	1.49E-10
F57G9.3	Hypothetical protein	8.82	8.70E-09
ins-19	Insulin related	8.75	1.06E-15
fbxb-3	Protein-protein interactions	8.70	1.64E-08
dsl-1	Postembryonic development	8.64	3.89E-09
fbxb-42	F-box B protein	8.56	2.92E-09
C40A11.6	Tumour necrosis factor	8.50	3.34E-09
fbxc-34	F-box C protein	8.48	2.30E-08
fbxb-13	F-box C protein	8.43	3.79E-09
F21H7.10	Hypothetical protein	8.26	3.00E-09
F45D11.5	Hypothetical protein	8.24	9.17E-08
C29G2.3	Hypothetical protein	8.24	8.70E-09
fbxb-10	Protein-protein interactions	8.17	1.07E-08
C17E4.19	Hypothetical protein	8.08	8.19E-08
lys-10	Lysozyme	8.04	1.63E-07
fbxb-44	Paraoxonase-like protein	8.02	5.21E-08
fbxb-40	Protein-protein interactions	8.01	1.59E-08
ins-2	Insulin related	7.93	1.93E-07
nlp-39	Neuropeptide-Like Protein	7.92	7.30E-08
fbxb-43	Protein-protein interactions	7.89	6.53E-08
fbxb-1	Protein-protein interactions	7.79	4.66E-07
Y73C8C.8	Zinc ion binding activity	7.76	6.00E-11
scl-12	SCP-extracellular protein	7.76	3.92E-03
C08F8.15	Hypothetical protein	7.76	4.53E-06
C24H12.3	Unknown	7.66	2.74E-07
T26E3.8	Hypothetical protein	7.62	1.36E-07
fbxb-108	Protein-protein interactions	7.59	6.89E-06
fbxb-22	Protein-protein interactions	7.59	6.36E-06
fbxb-65	Protein-protein interactions	7.56	1.46E-06

Table 36-continued

Name	Function	FC T16.vs.C16	FDR T16.vs.C16
Y45G5AM.5	Hypothetical protein	7.53	5.04E-07
fbxb-47	Protein-protein interactions	7.53	8.95E-05
fbxb-91	Protein-protein interactions	7.49	6.09E-12
fbxb-66	Protein-protein interactions	7.44	1.94E-16
clec-196	killer cell lectin-like receptor	7.38	3.02E-11
fbxc-22	Protein-protein interactions	7.37	2.31E-06
fbxb-95	Protein-protein interactions	7.35	1.52E-06
F38C2.7	RNA-binding protein	7.31	5.36E-04
col-121	Cuticle collagen	7.31	9.41E-19
C32B5.15	Protein-protein interactions	7.31	4.18E-06
T08B6.5	Nucleotide binding activity	7.30	2.02E-05
fbxb-19	Protein-protein interactions	7.30	1.83E-05
pes-10	Patterned expression site protein 10	7.29	4.14E-08
dct-13	RNA-binding protein	7.11	5.39E-05
tbx-43	DNA binding transcription	7.10	1.16E-03
21ur-14519	21U-RNA gene	7.09	2.89E-05
F36H5.4	Hypothetical protein	7.07	2.36E-05
hch-1	Epidermal growth factor	7.07	3.49E-19
sdz-11	Zygotic transcript	7.06	3.79E-04
fbxb-61	Protein-protein interactions	7.05	8.86E-06
T26E3.5	Protein-protein interactions	6.99	2.42E-05
R05D8.11	Hypothetical protein	6.98	7.04E-06
fbxb-52	Protein-protein interactions	6.97	6.59E-05
fbxb-46	Protein-protein interactions	6.97	4.51E-06
fbxb-50	Protein-protein interactions	6.95	6.62E-05
fbxb-60	Protein-protein interactions	6.94	7.95E-05
ZK816.4	Hypothetical protein	6.93	3.90E-05
fbxb-82	Protein-protein interactions	6.93	5.47E-05
ZK250.15	Unknown	6.88	7.11E-05
clec-266	Lectin domain family	6.87	1.86E-39
T02G6.11	Hypothetical protein	6.86	5.48E-05
fbxb-6	Protein-protein interactions	6.85	1.50E-05
F21D9.5	Unknown	6.85	1.55E-04
C54F6.5	Hypothetical protein	6.85	2.95E-05
sdz-31	Zygotic transcript	6.83	1.18E-04
fbxc-30	Protein-protein interactions	6.79	1.01E-03
F30A10.15	Hypothetical protein	6.75	5.48E-05
fbxc-29	Protein-protein interactions	6.72	1.86E-11
C32B5.7	Cysteine-peptidase activity	6.69	3.46E-04
ZK899.5	Hypothetical protein	6.66	1.35E-04
Y7A5A.6	Unknown	6.66	6.09E-05
ZC239.21	Hypothetical protein	6.65	7.53E-05
fbxb-33	Protein-protein interactions	6.64	5.33E-04
T24E12.13	Hypothetical protein	6.61	2.61E-04
fbxb-32	Protein-protein interactions	6.60	8.42E-05
fbxb-116	Protein-protein interactions	6.59	7.74E-05

Table 36-continued

Name	Function	FC T16.vs.C16	FDR T16.vs.C16
sdz-18	Zygotic transcript	6.59	1.67E-03
K08H2.2	Hypothetical protein	6.59	1.20E-07
C35D6.4	RNA-binding protein	6.58	1.99E-03
dsl-6	Putative transmembrane	6.58	4.91E-09
Y38H6A.3	Hypothetical protein	6.57	8.79E-04
fbxc-18	Protein-protein interactions	6.57	1.38E-11
fbxb-30	Protein-protein interactions	6.55	7.00E-04

Table 37. Major groups of up-regulated genes in *C. elegans* at 40 h of infection model compared with the non-infected 16 h control sample and their predicted function

Number of genes	Predicted function
26	Hypothetical proteins
12	Unknown
30	Protein-protein interactions
6	Zygotic transcripts

Table 38. Top 100 *C. elegans* genes up-regulated after 40 h of infection compared with the non-infected 16 h control sample

Name	Function	FC T40.vs.C16	FDR T40.vs.C16
ins-19	Insulin related	10.27	8.05E-21
sdz-4	Zygotic transcript	8.78	3.69E-10
fbxb-3	Protein-protein interactions	8.60	2.83E-08
lys-10	Lysozyme	8.57	9.22E-09
sdz-5	Zygotic transcript	8.52	2.11E-09
cyp-14A3	Cytochrome P450	8.35	5.35E-06
cpr-2	cysteine-type peptidase	8.34	3.84E-07
nlp-39	Neuropeptide-Like Protein	8.33	6.25E-09
C40A11.6	Tumour necrosis factor	8.22	1.78E-08
fbxc-34	Protein-protein interactions	8.19	1.05E-07
hnd-1	Transcription factor	8.14	2.64E-09
fbxb-10	Protein-protein interactions	8.13	1.42E-08
fbxb-44	Protein-protein interactions	8.11	3.10E-08
F57G9.3	Hypothetical protein	8.09	3.37E-07
T04C12.1	Unknown	8.05	6.64E-10
fbxb-42	Protein-protein interactions	8.01	7.20E-08
ZK899.6	Protein-protein interactions	7.99	4.01E-08
fbxb-13	Hypothetical protein	7.92	8.58E-08
dsl-1	Postembryonic development	7.89	2.56E-07
fbxb-108	Protein-protein interactions	7.87	2.19E-06
fbxb-40	Protein-protein interactions	7.84	4.89E-08
ins-2	ATP binding activity	7.82	3.55E-07
C17E4.19	Unknown	7.74	5.24E-07
fbxb-107	Hypothetical protein	7.72	6.03E-07
F45D11.5	Unknown	7.58	2.42E-06
C08F8.15	Transcription-binding Zn	7.50	1.30E-05
fbxb-22	Protein-protein interactions	7.35	1.73E-05
F36H5.4	Unknown	7.34	7.53E-06
C32B5.15	Hypothetical protein	7.33	3.87E-06
Y73C8C.8	Hypothetical protein	7.26	1.31E-09
Y57A10C.1	Cuticlin-Like	7.20	3.50E-04
Y45G5AM.5	Hypothetical protein	7.17	3.92E-06
fbxb-1	Protein-protein interactions	7.16	1.07E-05

Table 38-continued

Name	Function	FC T40.vs.C16	FDR T40.vs.C16
ZK822.9	Unknown	7.13	5.44E-04
tbx-43	DNA binding transcription	7.10	1.17E-03
C54F6.5	Hypothetical protein	7.07	1.10E-05
F38C2.7	RNA-binding protein	7.02	1.06E-03
T26E3.8	Hypothetical protein	7.02	4.91E-06
T08B6.5	Nucleotide binding activity	7.01	6.40E-05
fbxb-50	Protein-protein interactions	7.01	5.37E-05
fbxb-65	Protein-protein interactions	6.99	2.22E-05
F21H7.10	Hypothetical protein	6.99	6.73E-06
C24H12.9	Hypothetical protein	6.96	1.85E-05
F21D9.5	Hypothetical protein	6.92	1.21E-04
nspe-5	Nematode Specific Peptide	6.88	2.08E-04
clec-266	C-type lectin	6.69	6.39E-38
fbxb-95	Protein-protein interactions	6.66	5.17E-05
fbxb-82	Protein-protein interactions	6.65	1.63E-04
C24H12.3	Unknown	6.65	5.43E-05
ZK250.15	Unknown	6.65	1.82E-04
W08F4.13	Hypothetical protein	6.64	1.55E-04
fbxb-91	Hypothetical protein	6.62	3.38E-09
fbxc-22	Protein-protein interactions	6.60	9.00E-05
ZK816.4	Hypothetical protein	6.57	1.78E-04
fbxb-47	Protein-protein interactions	6.55	1.40E-03
fbxb-46	Protein-protein interactions	6.54	4.57E-05
C02B4.3	Unknown	6.52	1.94E-02
col-121	Cuticle collagen	6.51	3.33E-15
F21D9.11	Hypothetical protein	6.48	2.64E-03
C54C8.12	Hypothetical protein	6.47	2.20E-07
T02G6.11	Hypothetical protein	6.47	2.66E-04
Y7A5A.6	Unknown	6.45	1.53E-04
fbxb-62	Protein-protein interactions	6.45	7.96E-04
nhx-6	Sodium/proton exchanger	6.44	1.06E-04
C29G2.3	Hypothetical protein	6.44	1.38E-04
T24E12.13	Hypothetical protein	6.43	5.02E-04
Y46H3A.5	Hypothetical protein	6.43	2.84E-23
fbxb-18	Protein-protein interactions	6.43	1.02E-04
dct-13	RNA-binding protein	6.42	6.18E-04
sdz-11	Zygotic transcript	6.36	2.38E-03
fbxc-15	Protein-protein interactions	6.36	5.13E-04
fbxb-30	Protein-protein interactions	6.35	1.26E-03
dsl-6	Postembryonic development	6.35	1.89E-08
sdz-31	Zygotic transcript	6.30	7.99E-04
B0391.14	Unknown	6.29	8.62E-04
fbxb-61	Protein-protein interactions	6.27	3.15E-04
fbxc-30	Protein-protein interactions	6.26	3.63E-03
sdz-30	Zygotic transcript	6.26	1.28E-09
F35E12.2	Hypothetical protein	6.25	4.63E-04

Table 38-continued

Name	Function	FC T40.vs.C16	FDR T40.vs.C16
sdz-18	Zygotic transcript	6.23	3.95E-03
T04B8.2	Protein-protein interactions	6.20	3.10E-08
H25K10.4	Hypothetical protein	6.19	1.61E-02
F54F7.9	Hypothetical protein	6.15	1.86E-03
fbxb-116	Protein-protein interactions	6.14	5.30E-04
R05D8.11	Hypothetical protein	6.14	3.91E-04
fbxb-90	Protein-protein interactions	6.14	2.06E-06
F43C11.10	Unknown	6.13	2.87E-03

Table 39. Major groups of up-regulated genes in *C. elegans* at 40 h of infection model comparing to 16 h of infection sample and their predicted function

Number of genes	Predicted function
35	Unknown
27	Hypothetical proteins
8	Signaling receptors
4	Lectin proteins
3	Cytochrome P450
2	Collagen proteins

Table 40. Top 100 *C. elegans* genes up-regulated after 40 h of infection comparing with 16 h of infected sample

Name	Function	FC T40.vs.T16	FDR T40.vs.T16
C02E7.10	Hypothetical protein	7.05	2.30E-03
C54C8.2	Hypothetical protein	6.46	3.84E-03
C32H11.6	Hypothetical protein	6.18	2.99E-03
Y39G8B.7	Hypothetical protein	6.18	5.17E-03
T22B2.6	Hypothetical protein	6.18	2.85E-03
catp-2	Metal ion binding	6.16	1.76E-03
F14D7.10	Hypothetical protein	6.05	1.39E-03
B0563.10	Hypothetical protein	5.73	3.16E-02
C37A5.5	Hypothetical protein	5.70	4.67E-02
F07C6.6	Hypothetical protein	5.60	4.24E-03
T25G12.13	Oxidoreductase activity	5.42	6.04E-04
gba-2	Glucosyl ceramidase	5.29	9.11E-04
nhr-259	DNA transcription	5.21	1.70E-02
B0507.4	Hypothetical protein	5.20	1.34E-01
ZK563.10	Unknown	5.18	7.10E-02
nhx-6	Sodium/proton exchanger	5.17	8.90E-04
cyp-35A1	Cytochrome P450s	5.15	1.62E-03
R05A10.7	Hypothetical protein	5.10	5.98E-02
clec-206	C-type lectin	5.07	2.30E-03
F52B10.16	Unknown	4.99	1.10E-02
Y60A3A.23	Hypothetical protein	4.98	5.74E-02
nhr-74	Nuclear hormone receptor	4.90	8.13E-03
H06H21.36	Unknown	4.87	3.60E-02
C32H11.8	Hypothetical protein	4.85	8.88E-03
C32H11.9	Hypothetical protein	4.76	1.23E-02
C05E4.12	Hypothetical protein	4.74	2.93E-02
cyp-13A9	Cytochrome P450	4.73	3.51E-02
ZK1025.5	Unknown	4.73	4.00E-02
H39E23.3	Chloride intracellular channel	4.67	5.74E-07

Table 40-continued

Name	Function	FC T40.vs.T16	FDR T40.vs.T16
Y105C5B.1420	Unknown	4.62	5.24E-02
dao-4	Dauer Overexpression	4.61	1.38E-02
Y57G11C.41	Hypothetical protein	4.60	5.59E-02
W01B6.4	Hypothetical protein	4.59	5.67E-02
K02A2.9	Unknown	4.58	5.26E-02
C13A2.9	Hypothetical protein	4.57	6.21E-02
C17F3.4	Hypothetical protein	4.52	2.47E-01
F49C5.11	Unknown	4.48	1.25E-02
F36H1.17	Unknown	4.48	1.26E-01
nhr-222	Steroid hormone receptor	4.46	1.26E-01
sls-2.3	Unknown	4.46	1.21E-01
F17B5.4	Sulfotransferase activity	4.45	8.49E-02
C02H6.3	Hypothetical protein	4.45	1.69E-01
clec-247	C-type lectin	4.42	6.27E-02
C05A9.8	Unknown	4.42	1.46E-01
D1086.3	Hypothetical protein	4.24	4.46E-05
21ur-9901	Unknown	4.24	1.38E-01
E02C12.11	Transferase activity	4.23	8.30E-02
nhr-234	DNA binding transcription	4.23	1.33E-01
F48C5.3	Unknown	4.21	1.46E-01
math-29	Meprin-associated Traf	4.20	1.87E-01
C06E4.6	Oxidoreductase activity	4.17	4.16E-02
srz-15	Serpentine receptor	4.17	2.88E-01
srh-167	Serpentine receptor	4.13	3.07E-01
srz-55	Unknown	4.13	3.06E-01
col-163	Collagen	4.12	7.14E-03
B0495.21	Unknown	4.09	1.81E-01
K08B5.3	Unknown	4.09	1.81E-01
T26H10.6	Unknown	4.09	1.81E-01
cyp-13A11	Cytochrome P450 family	4.08	5.11E-03
T28C12.13	Unknown	4.07	9.41E-02
C34E7.5	Unknown	4.07	1.83E-01
F49F1.11	Lectin, galactoside-binding	4.07	1.73E-01
21ur-6449	Unknown	4.07	1.15E-01
F40F9.11	Unknown	4.07	1.30E-01
clec-2	C-type lectin	4.05	9.39E-02
C12C8.6	Unknown	4.05	1.11E-01
C52G5.7	Unknown	4.05	1.24E-01
ZK228.10	Hypothetical protein	4.04	2.10E-01
T07C12.10	Hypothetical protein	4.04	2.04E-01
R144.19	Hypothetical protein	4.04	1.73E-01
C26F1.13	Unknown	4.04	1.67E-01
hpa-2	Aging-associated changes	4.04	1.73E-01
C47E8.14	Unknown	4.03	1.81E-01

Table 40-continued

Name	Function	FC T40.vs.T16	FDR T40.vs.T16
F08H9.14	Unknown	4.03	3.12E-01
F33D4.13	Unknown	4.03	3.12E-01
F41G3.20	Unknown	4.03	1.21E-01
srh-188	Serpentine receptor	4.02	7.61E-02
F47B7.4	Unknown	4.02	1.31E-01
col-114	Collagen	4.02	8.21E-02
C27A12.11	Unknown	4.01	1.81E-01
T03G11.19	Unknown	4.01	1.29E-01
srg-38	Signaling receptor	4.00	1.73E-01
21ur-9425	Unknown	4.00	1.73E-01
cnc-1	Antimicrobial peptide	3.98	2.13E-01

Table 41. Major groups of down-regulated genes in *C. elegans* at 16 h of infection model comparing to 16 h control sample and their predicted function

Number of genes	Predicted function
39	Hypothetical protein
30	Cuticle collagen
7	Synthesize cuticle
6	Unknown

Table 42. Top 100 *C. elegans* genes down-regulated after 16 h of infection compared with the non-infected 16 h control sample

Name	Function	FC T16.vs.C16	FDR T16.vs.C16
C54C8.2	Hypothetical protein	-8.74	4.70E-05
Y39G8B.7	Hypothetical protein	-8.74	2.97E-05
T22B2.6	Hypothetical protein	-8.50	1.11E-05
C32H11.6	Hypothetical protein	-8.43	1.54E-05
nhr-74	Nuclear hormone receptor	-7.95	5.68E-06
cyp-13A9	Cytochrome P450s	-7.62	4.45E-05
clec-206	C-type lectin	-7.61	3.31E-06
ZK1025.5	Hypothetical protein	-7.39	1.53E-04
ZK1025.4	Hypothetical protein	-7.31	1.06E-05
F49C5.11	Hypothetical protein	-7.28	8.76E-06
dao-4	Aging adult Overexpression	-7.13	1.65E-04
nhr-259	Nuclear hormone receptor	-7.04	3.09E-04
T06G6.6	Hypothetical protein	-6.99	1.61E-06
pud-3	Protein Up-regulated Daf-2	-6.98	3.04E-06
bli-2	Cuticle collagen	-6.85	3.78E-03
nhr-113	Nuclear hormone receptor	-6.81	1.31E-04
clec-247	C-type lectin	-6.79	7.03E-04
pqn-57	Prion-like protein	-6.76	3.21E-06
W01B6.4	Hypothetical protein	-6.74	1.34E-03
C32H11.8	Hypothetical protein	-6.73	1.21E-04
bah-1	Nematode cuticle	-6.68	1.30E-04
col-49	Cuticle collagen	-6.65	5.21E-04
R166.8	Unknown	-6.65	4.32E-04
str-131	Seven TM Receptor	-6.64	7.11E-05
F07E5.7	Hypothetical protein	-6.62	4.83E-04
ZK1025.3	Hypothetical protein	-6.61	1.49E-04
E01G4.6	Hypothetical protein	-6.59	5.07E-04
D1014.6	Hypothetical protein	-6.52	1.44E-04
col-175	Cuticle collagen	-6.51	1.57E-03
C04G6.2	Hypothetical protein	-6.51	3.24E-07
rol-1	Synthesize adult cuticle	-6.50	2.68E-03
F26F2.10	Hypothetical protein	-6.46	1.20E-04
nhr-73	Seam cell development	-6.34	6.56E-05

Table 42-continued

Name	Function	FC T16.vs.C16	FDR T16.vs.C16
C33G8.13	Hypothetical protein	-6.30	5.25E-04
col-79	Cuticle collagen	-6.30	8.52E-04
dpy-4	Larval development	-6.30	2.77E-06
Y47G6A.15	Hypothetical protein	-6.29	8.50E-04
col-138	Cuticle collagen	-6.29	9.11E-04
F35F10.5	Hypothetical protein	-6.29	1.69E-06
col-71	Cuticle collagen	-6.29	2.24E-05
col-163	Cuticle collagen	-6.28	1.39E-06
sqt-1	Cuticle collagen	-6.27	1.24E-06
col-133	Constituent of cuticle	-6.27	3.29E-03
K02E11.10	Hypothetical protein	-6.26	1.70E-05
F28A10.4	Hypothetical protein	-6.05	6.66E-04
sqt-2	Cuticle collagen	-6.01	3.71E-07
col-137	Cuticle collagen	-6.01	4.16E-04
bli-6	Cuticular collagens	-6.00	1.24E-03
F31B9.4	Hypothetical protein	-6.00	1.88E-05
bli-1	Cuticular collagens	-5.99	5.94E-03
R05A10.7	Hypothetical protein	-5.97	2.34E-02
pqn-2	Prion-like protein	-5.96	8.98E-05
col-88	Collagen superfamily	-5.96	2.88E-03
ZC411.1	Hypothetical protein	-5.95	3.07E-03
col-176	Cuticle collagen	-5.95	1.99E-05
col-60	Collagen superfamily	-5.89	1.59E-03
col-149	Collagen superfamily	-5.88	1.88E-05
C11H1.5	Hypothetical protein	-5.88	3.59E-05
col-17	Cuticle collagen	-5.86	1.29E-04
C13A2.9	Hypothetical protein	-5.86	9.12E-03
dpy-13	Collagen superfamily	-5.84	1.24E-05
T08G3.13	Unknown	-5.84	9.58E-03
col-114	Cuticle collagen	-5.83	1.77E-03
C05A9.8	Unknown	-5.82	4.68E-02
F55C10.4	Hypothetical protein	-5.82	1.86E-04
col-145	Collagen superfamily	-5.82	9.44E-08
F57H12.6	Hypothetical protein	-5.82	4.53E-08
W05B10.3	Hypothetical protein	-5.81	1.87E-03
ZK1025.8	Sulfotransferase activity	-5.80	4.96E-04
Y46H3D.1	Hypothetical protein	-5.78	2.20E-03
abu-9	Transmembrane protein	-5.74	6.04E-04
lon-3	Cuticle collagen	-5.71	4.33E-06
D1086.3	Hypothetical protein	-5.71	4.84E-08
col-157	Cuticle collagen	-5.70	8.30E-08
ZK354.9	Hydrolase activity	-5.69	1.96E-02
dpy-20	BED zinc finger protein	-5.68	9.58E-04
Y47D7A.18	Hypothetical protein	-5.66	1.63E-02
col-41	Collagen superfamily	-5.64	2.04E-07

Table 42-continued

Name	Function	FC T16.vs.C16	FDR T16.vs.C16
C05G5.7	Hypothetical protein	-5.62	2.51E-07
Y47D7A.11	Hypothetical protein	-5.61	5.81E-03
rol-6	Cuticle collagen	-5.59	1.01E-05
F47B7.4	Unknown	-5.59	1.91E-02
ZK381.8	Hypothetical protein	-5.59	5.01E-03
col-38	Collagen superfamily	-5.57	2.21E-05
Y60A3A.23	Hypothetical protein	-5.55	3.03E-02
col-120	Constituent of cuticle	-5.51	8.68E-04
T06E4.12	Constituent of cuticle	-5.50	8.04E-05
Y51H7C.13	Hypothetical protein	-5.50	1.41E-03

Table 43. Major groups of down-regulated genes in *C. elegans* at 40 h of infection model comparing to 16 h control sample and their predicted function

Number of genes	Predicted function
46	Hypothetical protein
29	Unknown
13	Signaling receptor
4	Metabolic enzymes

Table 44. Top 100 *C. elegans* genes down-regulated after 40 h of infection compared with the non-infected 16 h control sample

Name	Function	FC T40.vs.C16	FDR T40.vs.C16
D1014.6	Hypothetical protein	-6.52	2.23E-04
F02C9.1	Hypothetical protein	-5.99	5.42E-03
clec-245	C-type lectin	-5.99	6.66E-04
Y4C6A.4	Hypothetical protein	-5.85	4.95E-03
Y54E5B.6	Hypothetical protein	-5.84	2.03E-02
Y102A5C.38	Unknown	-5.79	3.72E-03
Y73B3A.7	Hypothetical protein	-5.71	5.08E-02
ZK643.7	Hypothetical protein	-5.69	6.59E-03
Y47D7A.11	Hypothetical protein	-5.61	7.60E-03
F47F6.9	Hypothetical protein	-5.54	1.12E-02
Y47D7A.17	Hypothetical protein	-5.53	8.82E-03
Y32G9A.3	Transcript Isoform	-5.53	4.15E-02
ZK328.10	Unknown	-5.53	9.97E-03
srw-121	Serpentine receptor	-5.45	5.18E-03
Y116A8C.8	Hypothetical protein	-5.42	6.11E-03
Y116A8C.44	Hypothetical protein	-5.39	1.20E-02
srh-18	Serpentine receptor	-5.32	2.36E-02
R07H5.14	Hypothetical protein	-5.31	6.89E-02
AC3.16	Hypothetical protein	-5.31	1.12E-02
C29F5.5	Hypothetical protein	-5.30	5.24E-02
W04E12.10	Hypothetical protein	-5.24	7.13E-02
T26H10.2	Unknown	-5.24	2.45E-02
Y26D4A.12	Protease inhibitor	-5.24	2.61E-02
Y41C4A.21	Hypothetical protein	-5.24	1.53E-02
F49C5.10	Hypothetical protein	-5.20	4.18E-02
C06E8.7	Unknown	-5.20	1.67E-02
F56A6.7	Hypothetical protein	-5.20	1.37E-02
C24D10.2	Hypothetical protein	-5.19	8.53E-05
C13C4.8	Hypodermal syncytia	-5.15	3.00E-02
F20D12.7	Unknown	-5.09	4.74E-02
21ur-2482	Unknown	-5.08	8.39E-02
F14F7.t1	Hypothetical protein	-5.06	4.28E-02
Y47D9A.4	Unknown	-5.02	4.90E-02

Table 44-continued

Name	Function	FC T40.vs.C16	FDR T40.vs.C16
T16D1.1	hypothetical protein	-4.98	6.85E-02
21ur-11873	Hypothetical protein	-4.97	1.27E-01
str-24	Seven TM receptor	-4.97	8.32E-02
F35C12.5	Hypothetical protein	-4.95	8.65E-02
F47B10.4	Unknown	-4.93	7.62E-02
21ur-9702	Unknown	-4.93	7.93E-02
T05A8.6	Hypothetical protein	-4.93	1.04E-01
Y76B12C.10	Unknown	-4.93	1.04E-01
Y32F6B.6	Unknown	-4.92	1.70E-02
C08H9.8	Hypothetical protein	-4.88	5.84E-02
K01D12.18	Unknown	-4.88	1.91E-03
K02E11.8	RNA pseudouridylate synthase	-4.70	1.03E-01
srx-64	Serpentine receptor	-4.69	2.48E-01
21ur-11492	Unknown	-4.69	2.27E-01
R02D5.20	Hypothetical protein	-4.68	1.27E-01
21ur-6527	Hypothetical protein	-4.67	1.33E-01
C54H2.6	Unknown	-4.64	2.07E-01
cwp-2	Co-expressed With Polycystins	-4.64	2.10E-01
21ur-13468	Hypothetical protein	-4.64	2.16E-01
R01E6.5	Hypothetical protein	-4.62	3.32E-02
K11H12.10	Hypothetical protein	-4.62	8.70E-02
W02H5.14	Unknown	-4.61	4.84E-02
sra-34	Serpentine receptor	-4.60	1.29E-01
oac-41	Hypothetical protein	-4.60	9.10E-02
stdh-4	Putative steroid dehydrogenase	-4.59	1.41E-01
21ur-15088	Hypothetical protein	-4.57	1.42E-01
21ur-7901	Hypothetical protein	-4.55	1.32E-01
Y26G10.3	Unknown	-4.54	1.55E-01
srh-149	Serpentine receptor	-4.54	1.44E-01
W01B6.11	Hypothetical protein	-4.52	5.98E-02
Y116A8C.51	Hypothetical protein	-4.52	1.41E-01
C17B7.14	Hypothetical protein	-4.51	1.80E-01
anr-47	Unknown	-4.51	1.67E-01
C13A2.4	Hypothetical protein	-4.50	1.55E-01
F17E5.5	Unknown	-4.50	1.74E-01
nspd-2	Nematode Specific Peptide	-4.49	2.14E-03
R31.5	Unknown	-4.48	2.23E-01
F57B7.9	Unknown	-4.48	2.23E-01
srj-19	Serpentine receptor	-4.48	2.23E-01
21ur-11874	Hypothetical protein	-4.48	9.72E-02
F48G7.7	Unknown	-4.48	2.41E-01
21ur-11197	Hypothetical protein	-4.48	2.78E-01
ZC477.15	G-protein receptor	-4.48	2.70E-01

Table 44-continued

Name	Function	FC T40.vs.C16	FDR T40.vs.C16
F57B10.17	Hypothetical protein	-4.47	2.02E-01
str-188	Seven TM receptor	-4.45	8.30E-02
F56D6.12	Hypothetical protein	-4.43	1.13E-03
srx-21	Serpentine receptor	-4.42	1.81E-01
sru-25	Serpentine receptor	-4.41	1.70E-01
Y38F1A.13	Unknown	-4.40	8.17E-02
C36F7.25	Unknown	-4.39	8.09E-02
Y69A2AR.23	Hypothetical protein	-4.39	2.32E-03
T05B4.12	Hypothetical protein	-4.38	1.30E-01
C18H7.4	Tyrosine-protein kinase	-4.36	9.36E-03
clec-58	C-type lectin	-4.36	1.14E-01
F36H12.9	Unknown	-4.36	3.67E-04

Table 45. Major groups of down-regulated genes in *C. elegans* at 40 h of infection model comparing to 16 h of infection sample and their predicted function

Number of genes	Predicted function
50	Unknown
22	Hypothetical proteins
11	Signaling receptors
4	C-type lectin

Table 46. Top 100 *C. elegans* genes down-regulated after 40 h of infection compared with the 16 h of infection samples

Name	Function	FC T40.vs.T16	FDR T40.vs.T16
scl-12	SCP- extracellular protein	-7.76	2.41E-03
scl-13	SCP- extracellular protein	-6.46	3.43E-02
clec-134	C-type lectin	-6.20	1.19E-02
ZK381.57	Hypothetical protein	-5.45	7.20E-03
F49H6.3	Hypothetical protein	-5.33	6.03E-02
21ur-6059	Unknown	-5.05	6.78E-02
C06E8.7	Unknown	-5.04	2.06E-02
K01C8.11	Hypothetical protein	-4.99	2.37E-02
C06A12.8	Hypothetical protein	-4.98	6.98E-02
mir-44	Posttranscriptional regulation	-4.98	4.07E-02
21ur-15231	Unknown	-4.94	7.99E-02
Y60C6A.2	Hypothetical protein	-4.85	2.62E-02
F48A9.4	Unknown	-4.85	1.03E-01
F49C5.7	Hypothetical protein	-4.84	4.18E-02
F26C11.4	Unknown	-4.79	9.81E-02
ZK381.51	Hypothetical protein	-4.79	9.81E-02
clec-107	C-type lectin	-4.77	1.01E-01
T23D5.5	Unknown	-4.74	1.95E-02
Y4C6A.4	Hypothetical protein	-4.73	3.97E-02
B0334.17	Unknown	-4.72	1.07E-01
21ur-10725	Unknown	-4.70	5.66E-02
F02C9.1	Hypothetical protein	-4.70	4.88E-02
21ur-11874	Unknown	-4.66	7.62E-02
K10H10.9	Hypothetical protein	-4.66	1.34E-01
nhr-81	Nuclear Hormone Receptor	-4.64	1.29E-01
srh-279	Serpentine Receptor	-4.62	7.99E-02
21ur-6987	Unknown	-4.61	1.15E-01
srg-25	Signaling receptor	-4.60	7.43E-02
F10G8.11	Unknown	-4.60	9.09E-02
srg-24	Signaling receptor	-4.59	4.92E-02
sru-38	Neuronal cell	-4.57	1.16E-01
21ur-15668	Unknown	-4.52	2.43E-01
F11H8.6	Unknown	-4.50	1.41E-01

Table 46-continued

Name	Function	FC T40.vs.T16	FDR T40.vs.T16
srg-9	Signaling receptor	-4.49	1.08E-01
K01A2.12	Protease inhibitor	-4.48	1.16E-01
srh-22	Serpentine Receptor	-4.48	6.25E-02
Y26D4A.12	Protease inhibitor	-4.48	7.28E-02
K09C4.2	Transmembrane transporter	-4.48	7.34E-02
W09G10.t3	Unknown	-4.47	6.05E-02
F11D11.19	Unknown	-4.47	5.90E-02
F47G4.13	Unknown	-4.47	4.95E-02
Y116A8C.44	Hypothetical protein	-4.47	5.71E-02
F12F6.13	Unknown	-4.46	4.80E-02
F58H1.14	Unknown	-4.45	7.20E-02
math-11	Meprin-associated Traf	-4.32	1.44E-01
B0457.15	Unknown	-4.30	9.66E-02
C31H2.13	Unknown	-4.29	5.95E-02
21ur-14670	Unknown	-4.29	1.43E-01
C06E7.93	Unknown	-4.26	1.71E-01
srbc-39	Serpentine Receptor	-4.24	1.57E-01
21ur-14969	Unknown	-4.23	1.98E-01
gmd-2	GDP-mannose 4, dehydratase 2	-4.20	3.08E-01
F13A2.9	Hypothetical protein	-4.20	3.08E-01
K10F12.10	Hypothetical protein	-4.17	2.18E-01
21ur-213	Unknown	-4.17	2.18E-01
C08E8.11	Hypothetical protein	-4.17	1.77E-01
21ur-6455	Unknown	-4.17	1.64E-01
T27F7.7	Unknown	-4.16	1.80E-01
H22K11.9	Hypothetical protein	-4.16	1.55E-01
F52B10.11	Unknown	-4.16	1.30E-01
sre-21	Serpentine receptor	-4.15	1.92E-01
21ur-10289	Unknown	-4.15	1.92E-01
21ur-11377	Unknown	-4.15	1.12E-01
F32B4.6	Hypothetical protein	-4.14	1.92E-01
21ur-12971	Unknown	-4.14	2.19E-01
F57G12.7	Unknown	-4.14	1.11E-01
C35B1.3	Hypothetical protein	-4.13	1.73E-01
Y57E12B.7	Unknown	-4.13	1.11E-01
K04D7.8	Unknown	-4.13	1.10E-01
21ur-7340	Unknown	-4.13	1.80E-01
clec-253	C-type lectin	-4.13	3.16E-01
sri-42	Serpentine chemoreceptor	-4.13	8.33E-02
K06C4.7	Hypothetical protein	-4.12	1.39E-01
C04E7.1	Hypothetical protein	-4.12	8.24E-02
srab-13	Serpentine receptor	-4.11	2.04E-01
R13H4.11	Unknown	-4.11	8.10E-02
21ur-4829	Unknown	-4.11	1.80E-01

Table 46-continued

Name	Function	FC T40.vs.T16	FDR T40.vs.T16
T05A8.1	Hypothetical protein	-4.10	1.01E-01
M153.7	Unknown	-4.10	1.09E-01
anr-42	Hypothetical protein	-4.09	1.46E-01
Y116A8C.8	Unknown	-4.09	7.92E-02
clec-252	C-type lectin	-4.09	1.80E-01
Y57A10C.12	Unknown	-4.08	1.38E-01
F01E11.14	Unknown	-4.08	1.74E-01
C06E7.89	Unknown	-4.08	1.80E-01
F09F9.16	Unknown	-4.08	1.80E-01
F55A8.5	Unknown	-4.06	1.93E-01
21ur-13160	Unknown	-4.06	1.93E-01

Table 47. Top 100 up and down-regulated genes of *C. elegans* at different times during the infection model

Fold change ratio above and below the threshold of ±2, significantly differentially expressed genes were defined with FDR-adjusted P-value < 5%.

Name	Annotation	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
clec-266	Lectin domain	5.558241	6.87	1.86E-39	6.69	6.39E-38	-0.19	8.65E-01
cav-1	Embryogenesis	7.014377	6.00	3.57E-36	5.47	3.28E-31	-0.53	4.35E-01
C35E7.5	Unknown	6.962521	5.25	5.69E-30	4.76	1.83E-25	-0.50	4.80E-01
cht-1	Embryogenesis	6.907533	4.95	1.07E-28	4.46	4.58E-24	-0.50	4.67E-01
ttr-50	Keratin	5.198313	4.96	7.79E-25	4.37	5.07E-20	-0.60	3.80E-01
Y46H3A.5	Unknown	3.682821	6.24	5.77E-22	6.43	2.84E-23	0.20	8.85E-01
H29C22.1	Unknown	5.383573	3.89	3.60E-17	4.65	3.64E-23	0.76	2.10E-01
T24C4.2	Unknown	4.644558	4.73	1.78E-21	4.33	2.02E-18	-0.40	6.18E-01
K02B12.2	Unknown	4.069439	5.73	1.71E-21	5.07	2.02E-17	-0.66	4.02E-01
ins-19	Unknown	4.304115	8.75	1.06E-15	10.30	8.05E-21	1.53	8.90E-02
Y82E9BR	Unknown	4.36591	6.16	7.58E-20	5.64	3.60E-17	-0.52	5.96E-01
fbxc-51	Unknown	5.093167	5.26	4.39E-20	4.67	1.26E-16	-0.59	4.84E-01
hch-1	Epidermal growth	3.64563	7.07	3.49E-19	6.82	6.05E-18	-0.25	8.69E-01
cht-3	Constituent of cuticle	8.192169	2.82	2.45E-16	3.13	1.31E-19	0.31	6.22E-01
col-165	Collagen	6.017185	6.14	2.35E-20	4.58	7.81E-13	-1.57	2.21E-02
ces-2	Apoptosis	3.762572	5.95	4.44E-18	5.79	2.94E-17	-0.16	9.40E-01
cut-3	Unknown	4.945543	5.53	4.45E-20	4.22	8.68E-13	-1.31	4.12E-02
T05E12.3	Tumour necrosis factor	5.013994	3.93	8.75E-18	3.84	3.60E-17	-0.09	9.82E-01
col-121	Cuticle collagen	3.358249	7.31	9.41E-19	6.51	3.33E-15	-0.80	3.98E-01
W03F11.1	Eggshell synthesis	6.012397	3.42	7.99E-13	4.41	9.94E-20	0.98	9.10E-02
R09E10.5	Cell surface mucin	5.709052	3.08	5.46E-20	2.39	1.95E-12	-0.70	1.07E-01
sepa-1	Embryogenesis	4.457864	6.22	8.80E-18	5.76	9.35E-16	-0.45	6.61E-01

Table 47-continued

Name	Annotation	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
EEED8.15	Unknown	4.359395	4.77	8.80E-18	4.44	9.11E-16	-0.33	7.27E-01
dpy-14	Embryogenesis	6.979845	6.03	1.33E-17	5.15	7.69E-14	-0.88	3.49E-01
F10E9.12	Unknown	4.972003	3.68	1.53E-14	4.05	3.60E-17	0.37	6.32E-01
skpo-1	Unknown	7.148056	2.23	2.47E-15	2.33	1.20E-16	0.10	9.18E-01
T04G9.7	Unknown	5.985112	2.55	7.27E-14	2.88	2.70E-17	0.33	5.93E-01
ZK1053.4	Embryogenesis	4.583162	4.05	5.13E-16	3.83	1.65E-14	-0.22	8.43E-01
fbxb-66	Unknown	2.579982	7.44	1.94E-16	6.75	1.76E-13	-0.69	5.04E-01
F48E3.4	Serine-type endopeptidase	6.735226	2.26	8.86E-13	2.63	5.26E-17	0.37	4.80E-01
C08F1.10	Unknown	3.315901	6.44	5.77E-16	5.88	1.41E-13	-0.56	6.16E-01
C01G6.3	Unknown	7.276837	3.01	3.41E-15	2.95	1.04E-14	-0.05	1.00E+00
ifa-3	Constituent of cuticle	3.68215	4.90	8.67E-16	4.39	6.09E-13	-0.51	5.80E-01
ZK813.1	Unknown	7.010004	2.89	1.64E-15	2.66	2.81E-13	-0.23	7.41E-01
C44B7.5	Unknown	8.060923	2.50	1.32E-13	2.63	5.10E-15	0.14	8.88E-01
hsp-70	Unknown	7.43764	5.14	1.05E-12	5.67	8.59E-15	0.53	6.32E-01
Y110A2AL.4	Unknown	4.292483	3.56	1.47E-15	3.05	1.26E-11	-0.52	4.45E-01
his-1	H4 histone	8.002064	3.15	3.71E-17	2.02	1.09E-07	-1.13	8.17E-03
Y45F10C.2	Unknown	4.207213	2.90	2.59E-07	4.63	1.20E-16	1.73	2.25E-03
dpy-17	Constituent of cuticle	6.231157	5.62	1.10E-14	4.75	2.83E-11	-0.87	3.82E-01
Y82E9BR.1	Unknown	2.813558	5.35	4.25E-14	4.91	5.66E-12	-0.44	6.46E-01
col-74	Constituent cuticle	6.038969	6.02	5.98E-15	4.49	1.90E-09	-1.53	6.70E-02
zeel-1	Constituent of cuticle	2.597211	5.86	1.32E-13	5.52	3.93E-12	-0.34	7.92E-01
F44F1.6	Unknown	3.290147	5.28	1.02E-13	4.89	5.92E-12	-0.39	7.23E-01

Table 47-continued

Name	Annotation	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
T05H10.3	Unknown	3.085155	5.77	3.24E-14	4.71	7.05E-10	-1.06	2.20E-01
T21C9.13	Unknown	6.305646	2.33	7.94E-12	2.47	3.11E-13	0.14	8.75E-01
C34C6.7	Unknown	5.409637	2.96	1.44E-12	2.93	2.26E-12	-0.03	1.00E+00
jmjd-3.2	Putative histone H3	3.312448	5.03	2.56E-13	4.49	6.79E-11	-0.54	6.00E-01
B0281.5	Tumour necrosis factor	3.626724	4.68	8.20E-13	4.40	1.71E-11	-0.28	8.34E-01
ZK675.4	Unknown	3.726333	6.02	1.36E-13	4.96	8.71E-10	-1.06	2.70E-01
Y46G5A.7	Protein-protein interactio	3.123447	4.63	1.86E-13	3.97	3.69E-10	-0.66	4.51E-01
T04C12.1	Unknown	1.133165	8.93	2.18E-13	8.05	6.64E-10	-0.89	3.49E-01
pes-2.1	Embryogenesis	3.371121	5.72	2.45E-12	5.56	9.02E-12	-0.15	9.68E-01
pes-2.2	Ubiquitin-mediated protein degradation	3.252341	5.52	3.22E-12	5.32	1.86E-11	-0.20	9.32E-01
ttr-51	Unknown	7.14365	1.78	4.84E-09	2.20	1.69E-13	0.43	3.66E-01
hnd-1	Transcription factor	1.27087	9.10	1.66E-12	8.14	2.64E-09	-0.96	3.51E-01
his-38	Packaging of DNA	5.680307	3.31	4.30E-14	1.87	4.63E-05	-1.44	2.31E-03
dct-18	Germline tumour affecting	6.827166	-2.06	5.47E-11	-2.10	2.16E-11	-0.04	1.00E+00
F28D1.2	Unknown	3.381882	4.24	2.12E-12	3.59	3.14E-09	-0.65	4.53E-01
sdz-30	Embryogenesis	3.189749	7.13	5.20E-12	6.26	1.28E-09	-0.88	4.92E-01
D1054.10	Unknown	7.85303	2.50	4.20E-11	2.51	3.78E-11	0.01	1.00E+00
sdz-4	Embryogenesis	1.589117	9.25	1.53E-11	8.78	3.69E-10	-0.47	7.42E-01
fbxb-41	Protein-protein interactio	2.173568	6.48	5.68E-12	5.66	3.14E-09	-0.82	4.51E-01

Table 47-continued

Name	Annotation	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
fbxb-91	Protein-protein interactions	1.797952	7.49	6.09E-12	6.62	3.38E-09	-0.87	4.32E-01
hsp-16.11	Heat shock proteins	7.019073	3.98	5.73E-10	4.28	3.10E-11	0.30	8.29E-01
ilys-2	Lysozyme activity	4.027286	5.92	4.40E-12	4.50	8.37E-08	-1.41	1.34E-01
clec-196	C-type lectin	2.639523	7.38	3.02E-11	6.82	9.56E-10	-0.56	6.83E-01
epg-2	Autophagy	7.326239	2.65	5.68E-12	0.03	1.04E-08	2.22	5.02E-01
fbxc-29	Protein-protein interactions	1.927203	6.72	1.86E-11	0.07	3.14E-09	6.02	5.64E-01
T27A1.3	Protein-protein interactions	1.65513	6.47	1.27E-11	5.70	6.25E-09	-0.77	4.70E-01
skr-7	Embryogenesis	2.777746	4.58	2.95E-10	4.72	7.30E-11	0.15	9.64E-01
cutl-2	Cuticlin-like protein	2.252654	5.51	3.10E-11	5.06	1.41E-09	-0.45	7.01E-01
tbh-1	Tyramine Hydroxylase	7.093627	1.77	2.16E-09	1.96	2.16E-11	0.19	7.42E-01
D1054.11	Unknown	8.485427	2.37	6.15E-10	2.50	5.69E-11	0.13	9.18E-01
inx-3	Embryogenesis	6.733508	2.82	8.11E-11	2.72	4.10E-10	-0.10	9.66E-01
vet-6	Embryogenesis	4.164211	6.03	3.33E-11	5.33	2.93E-09	-0.69	5.99E-01
clec-88	Lectin like	7.594933	2.22	6.02E-11	2.09	8.95E-10	-0.13	8.96E-01
Y73C8C.8	Zinc ion binding	2.192917	7.76	6.00E-11	7.26	1.31E-09	-0.50	7.33E-01
F44E5.4	Heat shock proteins	9.782001	4.50	3.24E-10	4.57	1.89E-10	0.06	1.00E+00
arrd-1	AR Restin Domain protein	3.141409	5.57	1.77E-10	5.38	7.04E-10	-0.19	9.53E-01
hsp-16.48	Embryogenesis/ Constituent of cuticle	6.52384	3.42	2.74E-09	3.75	7.30E-11	0.34	7.59E-01
sdz-5	Embryogenesis	1.324043	8.98	1.00E-10	8.52	2.11E-09	-0.47	7.48E-01
C08F1.6	Unknown	2.663316	5.91	6.03E-11	5.26	5.71E-09	-0.65	6.08E-01
F44E5.5	Heat shock proteins	9.874649	4.51	5.75E-10	4.59	3.13E-10	0.07	1.00E+00

Table 47-continued

Name	Annotation	Log CPM	FC T16.vs.C16	FDR T16.vs.C16	FC T40.vs.C16	FDR T40.vs.C16	FC T40.vs.T16	FDR T40.vs.T16
nhr-17	transcriptional regulators	6.205633	3.06	1.36E-08	3.57	3.78E-11	0.51	5.71E-01
Y41D4B.26	Unknown	2.199906	6.02	1.06E-10	5.53	3.80E-09	-0.49	6.93E-01
F54D5.5	Unknown	5.592415	2.74	1.56E-10	2.57	2.05E-09	-0.17	8.88E-01
T25E12.6	Unknown	2.958206	4.61	8.85E-11	4.14	6.45E-09	-0.47	6.41E-01
hsp-16.49	Heat shock proteins	6.479764	3.57	3.51E-09	3.90	1.38E-10	0.33	7.86E-01
pqn-73	Prion-like protein	4.773582	2.82	1.92E-11	2.21	2.59E-07	-0.62	3.12E-01
fmo-2	Oxidative metabolism	6.793298	2.71	3.02E-06	3.93	1.14E-11	1.22	7.51E-02
cyp-35A3	Encodes cytochrome P450s	5.971488	-4.82	9.55E-12	-1.97	0.007816	2.85	5.17E-05
cyp-35A2	Encodes cytochrome P450s	7.178269	-2.47	2.17E-12	-1.26	0.001063	1.21	1.93E-03
col-10	Cuticle collagen	7.346566	-3.01	2.18E-08	-3.48	1.06E-10	-0.47	6.13E-01
ZK899.6	Unknown	1.056743	8.86	1.49E-10	7.99	4.01E-08	-0.87	4.67E-01
C28G1.4	Zinc finger protein	2.472117	5.17	6.51E-10	4.98	2.64E-09	-0.18	9.48E-01

5.3 Discussion

The expression of *S. aureus* virulence genes was measured in different types of studies, but few have investigated the overall patterns of lipoprotein genes expression in *in vitro* and *in vivo* conditions in the same time, this study focused on lipoprotein genes expression levels in some *S. aureus* strains. Furthermore the transcriptome analysis in this study provides the first analysis of *S. aureus* lipoprotein proteomics in response to *C. elegans* infection challenge and enables an investigation of the roles lipoproteins may play in pathogenicity during infection.

5.3.1 Investigation the *S. aureus* lipoprotein genes expression *in-vitro* by qRT-PCR

Many studies have used the qRT-PCR technique to detect the presence of particular *S. aureus* lipoprotein genes (Hashimoto *et al.*, 2006a). Expression levels of the examined lipoprotein genes were observed in each growth condition were in different pattern of up or down-regulation of expression. The examined lipoprotein genes were selected at the beginning of this study did not have well-defined functions and in sequence annotations were predicted as putative lipoprotein. Some transcriptional profile differences were observed despite the identical growth conditions used. These variations in expression suggest that lipoprotein gene roles may not be similar in all strains.

Notably, these 5 lipoproteins were down-regulated during infection of *C. elegans*, apart from *SAR0740* gene which was up-regulated in the early time of *C. elegans* infection with *S. aureus* MRSA252. *S. aureus* transcriptional analyses during the growth in *in vitro* conditions did not correlate with gene expression in *in vivo* mammalian models or in human infections, therefore investigations to characterize *S. aureus* virulence genes regulation *in vivo* are preferred (Pragman and Schlievert, 2004).

In the current study, two approaches were used to detect lipoprotein gene expression levels, the first was qRT-PCR and the second was RNA sequencing. The plan for this study was to evaluate the majority of lipoprotein genes expression patterns using the qRT-PCR but using qRT PCR was replaced by the whole RNA sequencing due to the high-throughput of RNA sequencing and also their ability to determine gene expression level changes during the infection course.

5.3.2 Transcriptome of *S. aureus* 8325-4 lipoprotein genes *in vitro* under non-infection condition

The ratio of fold up-regulation for 5 cell wall-associated transporter binding lipoprotein transcripts strongly induced during the early stationary phase including peptide ABC transporter and ABC-type metal ion transporter had up-regulated expression level during the exponential growth phase. These types of transporter lipoproteins are involved in the uptake of different nutrients and antimicrobial resistance through the cytoplasmic membrane, also playing a role in peptidoglycan synthesis and maintenance and virulence of pathogenic bacteria, therefore, mutation of a few lipoprotein genes does not affect high-affinity different nutrients uptake but mutants are attenuated in virulence (Schmaler *et al.*, 2009).

Lipoprotein genes for iron uptake are over-expressed under iron deficiency conditions *in vitro* and *in vivo* indicating their roles in bacterial growth and survival (Allard *et al.*, 2006). Transcriptome of *S. aureus* 8325-4 at exponential growth phase had two up-regulated lipoprotein transcripts for cell-wall binding lipoprotein *SAR1066* and hypothetical lipoprotein *SAR0390* with different expression patterns observed during this phase, transcription of cell-wall binding lipoprotein *SAR1066* up-regulation level was continued in the stationary growth phase. Genome-wide transcriptional profiling of *S. aureus* 8325-4 has identified ~100 transcripts been up-regulated in response to the cell-wall-active antibiotics, this up-regulation in gene expression suggested the bacterial attempt to defend against the antibiotics activities (Utaida *et al.*, 2003). Proteomics study on *S. aureus* MRSA USA300 based on murine infection model revealed that the majority of surface proteome post- infection is proteins involved in nutrient acquisition and iron uptake including the lipoproteins of ABC-type transport systems (Diep *et al.*, 2014b). The oligo-peptide permease lipoprotein (OppA) *SAR0953* which is assist in protein folding (Stoll *et al.*, 2005), was among the up-regulated lipoproteins with ~ 6 fold change in the early time of growth.

The genes found to be down-regulated in all three time-points of evaluation in this study were not significant, as a result of no stress was applied on the bacterial growth conditions the bacterial growth curve was in the normal pattern and at the stationary phase the rate of growth decreased because of the lack of nutrients in the medium. The results indicate that up-regulated lipoproteins under non-infection condition were not associated with the *C.*

elegans defence functions, these observations suggest that these lipoprotein genes are not required for *S. aureus* pathogenesis and survival during infection.

5.3.3 *C. elegans* life span assay

This assay demonstrates that *S. aureus* MRSA252 exposed to *C. elegans* able to complete killing of *C. elegans* at 3 d in solid plates. However, earlier report by JebaMercy *et al*. indicated that 8 h of exposure to penicillin-resistant *S. aureus* ATCC 11632 was sufficient to kill *C. elegans*, also the accumulation of *S. aureus* in *C. elegans* intestine was found to be significantly increased from 24 to 48 h of exposures (JebaMercy *et al.*, 2011). In this study, the *C. elegans* exposed to *S. aureus* MRSA252 displayed abnormal movement and pharyngeal pumping, however the nematodes became immobile and the mortality rate was significant high as worms died after ~18 h of infection. Sifri *et al*., have examined the accumulation and proliferation of *S. aureus* inside the intestinal lumen of *C. elegans* by feeding worms with a lawn of *S. aureus* containing a shuttle vector which expresses *Aequorea victoria* green fluorescent protein (GFP), after 1 d of feeding the confocal fluorescence microscopy showed a considerable number of fluorescent cocci in the distended intestinal lumen (Sifri *et al.*, 2003). The infected nematodes become excessively weak to lay eggs normally then eggs of a gravid hermaphrodite hatched inside the worms and the "bag-of-worms" phenotype was observed.

5.3.4 Changes in lipoprotein transcriptome of *S. aureus* MRSA252 caused by *C. elegans* infection model

The lipoprotein gene up-or down-regulation in this experiment was basis on the fold change in expression between two *in vivo* growth conditions. A total of 62 lipoprotein transcripts were identified, these lipoproteins were divided into two main groups (up-or down-regulated) based on their response to the *C. elegans* infection and how they responded to the infection at different time-points.

During *S. aureus* infection the data of *C. elegans* transcriptome showed that different genes induced by infection were affected from one stage to another. The correlation between bacterial virulence factors in invertebrate and mammals has been reported as different genes have been shown to be involved in *in vivo* pathogenicity including many surface proteins such as capsular polysaccharide genes (Cheung *et al.*, 2004), in this study the expression of capsular polysaccharide synthesis enzymes of *S. aureus* were highly

significant as 16 transcripts of these enzymes had up-regulation fold changes at the early post-infection stage. *S. aureus* global virulence regulators *agr* and *sarA* promote production of many extracellular and cell wall-associated proteins and have been found to have a role in nematocidal activity (Sifri *et al*., 2003).

In another experiment of murine systemic infection with *S. aureus* USA300 MRSA performed by Diep *et al*., 15 putative lipoproteins genes analogous to *S. aureus* MRSA252 *SAR2457, SAR0118, SAR0761, SAR0872, SAR0390, SAR0790, SAR0839, SAR1288, SAR2545, SAR2470, SAR2368, SAR2546, SAR2563, SAR1932, SAR0641* were involved and showed high up-regulation during the course of the murine infection, only one of these lipoproteins *SAR2457* was involved in the *C. elegans* infection model and displayed significant response, while the other 14 genes had a different pattern of expression during the *C. elegans* infection, 6 transcripts of surface-associated lipoproteins *SAR2470, SAR2368, SAR2563, SAR2546, SAR1932* and *SAR0641* which are involved mainly in nutrients uptake and metal ions acquisition were down-regulated at the late stage of infection. However, a search on the *S. aureus* MRSA252 genome for the lipobox motif confirmed that *SAR2545* and *SAR2563* do not contains the lipobox sequence in their first 35 amino acids and in particular the invariant C was missing, these proteins were wrongly annotated as lipoproteins.

Seven lipoproteins genes including *SAR0230* putative extracellular solute-binding lipoprotein, *SAR2363* putative molybdate-binding lipoprotein precursor, *SAR2536* glycine betaine/ carnitine/ binding lipoprotein precursor, *SAR1878* and *SAR0340* putative lipoproteins, *SAR1402* phosphate-binding lipoprotein and *SAR1932* peptidyl-prolyl cis-isomerase have showed low expression during the early point of evaluation at 16 h, these lipoproteins remains to be in low expressed at the late stage of *C. elegans* infection model except *SAR1402* which showed ~ 5 fold up-regulation. The analogues of *SAR2368 SAR2470, SAR1932, SAR2546* and *SAR0641* in the murine infection model had over-expression levels, but in this study these genes were significantly down-expressed to suggest that they had no functions in *C. elegans* infection conditions.

In the late phase of infection 16 lipoproteins including 10 genes involved in uptake of metal ions and ABC transporter extracellular binding were identified to be down-regulated, these proteomics data indicate that the main down-expressed lipoprotein genes of MRSA252 during the *C. elegans* infection were lipoproteins that are related to nutrient

acquisition. MntC ABC transporter extracellular binding lipoprotein SAR0641 a part of MntABC operon is involved in the manganese acquisition was among the transcripts with a low production level during the late infection stage. SAR1932 peptidyl-prolyl cis-/trans-isomerase (PrsA) this lipoprotein involved in protein folding including formation of the cell wall and toxins (Heikkinen *et al.*, 2009), *PrsA* transcript had repressed level at the all points of evaluation in the infection model, the reason for this down-expression remain unclear. *SAR2104* putative lipoprotein was not found in clinical isolate RN4282 strain, this gene was inactive during the *C. elegans* infection model experiment as shown zero read transcript in the all period of infections.

Nematodes do not have an adaptive immune system or mobile phagocytic cells, and relies mainly on the innate immune system, the intestinal epithelial cells have a significant to defend the worms against pathogenic bacteria. While, vertebrates have more sophisticated immunity systems that participate the innate responses, phagocytic cells and antimicrobial peptides. For this reasons the comparison between the innate immunity response of *C. elegans* against *S. aureus* infection will not be similar to the vertebrates immunity response to an invading pathogens.

5.3.5 *C. elegans* transcriptional changes induced by infection with MRSA252

Although not part of the main study, the innate immune response of *C. elegans* is an important model for host defences against pathogens such as *S. aureus* which are capable of infecting and killing *C. elegans*. The detailed mechanisms of immune responses of *C. elegans* against invading pathogens are generally unknown, the general changes in genes transcription of *C. elegans* after infection with *S. aureus* showed that many genes were involved in the immune response of nematode in different levels to infection. This study determined lists of ~18,000 transcripts that were up or downregulated in various infection stages in response to bacterial pathogenicity.

The expression of cuticle collagen genes which are involved in many vital roles in the *C. elegans* body structure and secretions. Collagen deficiency can result in a series of defects from abnormal morphology of adults and larval death (Page and Johnstone, 2007). Cuticle collagen transcripts in early phase of infection had been down regulated as many of cuticle collagen transcripts showed low level of expression, the explanation for this down-expression are not clear, but the *C. elegans* in general at this stage of growth does not

require abundant cuticle collagen production as the full size of adult nematodes are reached after 3 d of growth.

Another observation from the RNA-seq data was the level of expression among the large gene family of C-type lectins in *C. elegans,* many suggestions that C-type lectins play a role in immune recognition of pathogens and general stress response functions (Schulenburg *et al.*, 2008). C-type lectin family genes expression in *C. elegans* during the *S. aureus* exposures showed up-regulation in sequentially post-infected conditions (JebaMercy and Balamurugan, 2012). There were also 13 F-box containing proteins were up-regulated in different phases of infection, to suggested ubiquitin ligases involvement in host defense (Thomas, 2006). One of the important innate immune system proteins were found among the up-regulated transcripts is the lysozyme genes, infected *C. elegans* with bacterial pathogens subjected to many changes in gene regulation of antimicrobial peptides and proteins include the lysozyme encoding genes (Mallo *et al.*, 2002).

The transcripts of G-protein-coupled receptor that sense molecules outside the cell and activate the signal transduction pathways were over-expressed at many stages of infection. The genetic features of sensory neuron specific G-proteins shows that olfaction, nociception and pheromone responses in *C. elegans* are controlled by G protein-coupled receptor (Jansen *et al.*, 2002). On the other hand, RNA-seq data showed that signaling receptor during the infection period were highly expressed, the chemoreceptor families of *C. elegans* mediated by members of seven-transmembrane G-protein-coupled receptor including the large family of *str* genes, the prospective function of these receptors are based on transgene expression patterns in the known pairs of chemosensory neurons (Robertson and Thomas, 2006).

Transcription factors (sequence-specific DNA-binding factors) are involved in *C. elegans* embryogenesis and many of these genes function as a key regulators of early blastomere divisions, but how these genes are up-regulated to regulated zygotic developments is not understandable (Maduro and Rothman, 2002). The early mature hermaphrodite nematodes used in this experiment have a large number of cells and contain eggs/embryos under development stages this would contribute to the RNA-seq results as many embryogenesis transcripts will be observed.

Also, a set of the apoptotic germline cells genes was induced by the *S. aureus* infection, these genes is a part of the oogenesis program in *C. elegans*, the physiological germline apoptosis occurs in the absence of any known stress and is initiated by the core apoptotic machinery, besides it can be initiated by infection with pathogenic bacteria (Lettre and Hengartner, 2006). The bacteria invasion led to an up-regulation of 32 F-box B protein genes, the physiological function of F-box B protein control apoptotic cell death and the stability of some antiapoptotic proteins (Chiorazzi *et al.*, 2013). Over-expression of insulin related gene (*ins*) is remains unclear whether this effect was directly related to the early infection phase, the overexpression of *ins* genes were reported to enhance dauer arrest in certain stages (Pierce *et al.*, 2001).

Cysteine protease genes in RNA-seq results of *C. elegans* were up-regulated after 16 h of infection, cysteine protease has reported a role in innate immunity against *S. aureus* which act in the degradation of proteins and antigen processing (Irazoqui *et al.*, 2010a). Several genes of prion-like Q/N proteins known to be involved in *C. elegans* stress response comprised were down-regulated in the late stage of bacterial exposure (Golden and Melov, 2007). Transcriptome results revealed some cellular biological effects of *S. aureus* infection on the host innate immunity including detoxification factors cytochrome P450 genes which are expected to function in the detoxification of xeniobiotics (Werck-Reichhart and Feyereisen, 2000), these genes displayed over-expression at different time points of infection.

This experiment showed significant down-regulation of ~ 3000 transcripts at 40 h post-infection influenced the physiological and behavioural activities of the nematodes with the majority of metabolic enzymes and signalling receptor transcripts as the nematodes showed visible signs of illness and starvation. Nine metabolic enzyme genes known to be involved in glycoprotein biosynthesis in *C. elegans* were down-regulated in the late stage of bacterial infection experiment. Microarrays transcription profiling differences for the dauer state/non-dauer state and after feeding of starved L1 *C. elegans* showed changes in expression levels for large number of genes involved in the metabolism process (Wang and Kim, 2003).

5.3.6 Post-infection differentially-regulated genes of *S. aureus* MRSA252 in *C. elegans* infection model

Investigation of time-based changes of genes expression using an RNA-seq approach to identified *S. aureus* response to *C. elegans* infection has identified markedly different response of *S. aureus* to host innate immune mechanisms, the overall transcriptomic data showed that 25% of differentially transcribed genes observed in the post-infection course, after 40 h of infection course 1274 identified transcripts known to be significantly differentially-expressed in response to the *C. elegans* challenge, 799 transcripts were differentially-expressed at 16 h post-infection including 349 up-regulated and 450 down-regulated, the number of transcripts up-or-down expressed that report as differentially expressed shown similar number of transcripts at the both stages of infection, the rapid dynamics in over-expression of some genes after short time of *C. elegans* challenge and suggesting their roles in pathogenicity and virulence of *S. aureus*.

Differentially expressed transcripts in both 16 and 40 h post-infection stages showed 118 up-regulated and 189 down-regulated transcripts. The large-scale transcriptome of *S. aureus* pathogenicity are highly dependent on the infection conditions to show the interaction between host response and pathogen. The continuation of bacterial exposure has revealed 307 genes were conserved in expression post-infection indicating their importance for host pathogenesis.

5.3.7 Changes in expressed genes of nematodes during the infection of *C. elegans*

To understand the nature of *C. elegans* response to *S. aureus* infection, differentially expressed genes (up-or-down) displays the impact of infection on *C. elegans* and to show genes involved during infection. Transcriptomic data showed that 14% of differentially transcribed genes were observed in different abundance at the time-course of infection. The differentially expressed genes of *C. elegans* that observed in the early exposure time-point have showed that transcripts number were almost similar as 1853 up-expressed genes than 1776 down-expressed genes with average of 1:1 expressed in response to the *S. aureus* challenge. However, *C. elegans* transcripts at 40 h were quite different; there were more up-expressed genes than down-expressed genes in the late stage of *S. aureus* exposure as showed differentially expressed 1880 genes were up-regulated and 1311 genes down-regulated, those genes showed lower expression level in this stage were less than the over-expressed genes.

In order to find out the genes expression profile comparisons between the 16 and 40 h post-infection to evaluate the changes in genes expression, 231 transcripts were observed to be expressed in both conditions and may play a part in the immunity response, while 53 down-regulated transcripts to suggests that their expression may not be involved in host resistance against bacterial infection.

Chapter six
General summary and discussion

6. General summary and discussion

The main objective of this study was to try to clarify the number and genetics of *S. aureus* lipoproteins and to determine whether the genes were expressed under different conditions. The results generally extended the information of the molecular variation of lipoprotein genes by examining and comparing the genetic background of lipoproteins in some *S. aureus* strains that are involved in human and animal infection. The BLAST searches and PCR results revealed certain variation in the lipoproteins among the examined strains, six genes were identified in only few strains also a number of lipoprotein genes in RN4282 and T1 strains were not detected in MRSA252. The majority of mutations in the lipoprotein genes were represented within a limited number of highly variable genes to reveal significant variation among *S. aureus* strains, there were 11 highly variable genes with considerable value of non-synonymous residues. Whereas, almost half of the genes showed low nucleotide diversity, these genes were fairly invariant in examined *S. aureus* strains. The numbers of lipoprotein genes were variant in these strains as the predictive algorithms rules used to predict the unique N-terminal lipid modification, *N*-acyl-*S*-diacylglyceryl-Cys in the complete genome of *S. aureus* have revealed different number of lipoprotein genes. One confounding factor was that lipoproteins were sometimes wrongly annotated of genes in the gene prediction and functions in different genetic sequence database and collection of publicly available DNA sequences which could lead to incorrect analysis. Some assigned genes sequences in different databases were annotated as lipoprotein but were missing the C-region end of signal peptides with Cysteine (+1 position) which is the lipid-modification site. *SAR2104* (putative lipoprotein) was not detected in RN4282 *S. aureus* strain, this gene was inactive during the *C. elegans* infection model experiment as shown RNA zero read for the all period of infections. In the PCR approach to detect the lipoprotein genes of *S. aureus* were limited to 50 genes at the beginning of this study, while the RNA sequencing analysis has identified 62 lipoprotein transcripts in total, the extra 12 lipoprotein genes need to be investigated for their genetic diversity and the number of mutations. Lipoprotein genes sequence detected variable SNPs between *S. aureus* strains including 10 genes with intense nucleotides changes leading to excessive amino acids variations, the phylogenetic trees were more effective by the genes with a highly number of variable SNPs. A detailed proteomic profiling of *S. aureus* lipoproteins is important to confirm the translation of their mRNA's and to determine their potential roles as virulence factors and antigens. This information could help to identify

the differences and similarities of lipoprotein constituents and provide helpful information to develop a vaccine and/or treatment of *S. aureus* infections. Proteomic analysis using the application of gel-free proteomic technique to identify the lipoproteins within *S. aureus* MRSA252 identified 38 lipoproteins that were expressed in the non-infection condition representing approx. two-thirds of the *S. aureus* MRSA252 lipoproteins. Meanwhile, in the *S. aureus*-mediated infections with *C. elegans* showed involvement of 62 lipoprotein genes, identification of some lipoproteins in the infection condition that were not detected in the proteomic profile in non-infected samples suggest that these lipoproteins might be associated with bacterial virulence. Also, detection of transcripts of some lipoproteins does not prove they are translated and synthesized into final active lipoprotein form, as several steps in the protein expression process and many factors can affect the processes of protein translation and post-translational modification. An analytical study to investigate the relation of transcription, translation, and cellular protein turnover in the yeast *Saccharomyces cerevisiae* by comparing protein to mRNA ratios and the translational activity revealed a clear correlation between protein abundance and mRNA are significantly related to biological functional e.g. cellular metabolism and energy (Beyer *et al.*, 2004).

S. aureus infection of *C. elegans* induced strong transcriptional immunity host response to support the view that *S. aureus* infection stimulates the innate immune system of *C. elegans* to regulating and produce more antimicrobial products during the invasion, confirming previous studies on knowledge of the molecular biology of the *C. elegans* genes that are involved in the transcriptional response to *S. aureus* infection (Irazoqui *et al.*, 2010a). The up-or down-regulated changes in genes transcription in *C. elegans* after infection with *S. aureus* shown expression of immune-related genes and virulence factors that were involved in the post-infection virulence of the nematode, these genes needs to be subjected to further investigation as many of them were determined to be important in mammalian pathogenesis (Irazoqui *et al.*, 2010b). In the other hand, the *C. elegans* infection model is not directly comparable to animals and virulence factors also vary between different species and tissues, therefore the pathogenesis models are variable in the assessment of the host response and bacterial virulence factors. Also, the time of RNA extraction and the age of worms are very critical to detect and evaluated some important transcripts level, *C. elegans* has a short generation time which is 3 days cycle at 25°C to produce progeny, RNA isolation from mature worms may contain some RNA of

developed eggs and new progenies that may hatched inside the worms. High level of some zygotic and embryonic transcripts have been detected during the early stages of *C. elegans* infection which is more likely linked to development of eggs into infected *C. elegans* as a result of egg-laying behaviour was affected by *S. aureus* infection (Gardner *et al.*, 2013). However, funding was only available to analyse samples from three times of infection. The results demonstrate that *S. aureus* MRSA252 was able to infect and kill *C. elegans* under the examined conditions of *in vitro* infection model, the *C. elegans* infection experiment showed prospective interactions of *S. aureus* antigen expression and antibody response to antigens in early and late stages of infection, but the mechanisms of host-pathogen interaction still to be clarified. Transcriptome observation showed that the *C. elegans* host response post-infection shares important factors with the mammalians innate immunity response to suggest their importance in human innate responses.

Some lipoproteins shown to be involved in experimental infection of mice experiment done by Diep *et al*. (Diep et al., 2014a) were not involved in pathogenesis in *C. elegans* in this study, while putative membrane protein *SAR0706* and phosphate-binding lipoprotein *SAR1402* lipoprotein genes were significantly up-regulated in *C. elegans* infection experiment and identified as virulence factors, these lipoproteins promising to be an effective component of vaccine candidates against *S. aureus* but require further investigation. Approx. 60 lipoprotein genes expressed in the clinically important *S. aureus* MRSA252 some of them were known as factors of pathogenicity, transcriptional profiling of *S. aureus* post-infection showed that *C. elegans* infection model elicits changes in expression of the majority of the total genome of ~ 2,600 genes. The up-regulation of capsule synthesis proteins during infection stages was the most significant virulence factor that was related to *S. aureus* infection.

Label-free shotgun proteomics techniques based on the identification of peptides by LC-MS/MS has provided a high dynamic range of lipoproteins to cover almost two-thirds of the *S. aureus* MRSA252 predicted lipoproteins, while, the detected lipoprotein transcripts in RNA-seq of *C. elegans* infection experiment identified 62 transcripts, the two methods have been performed in two different conditions and showed a higher number of lipoprotein transcripts that may not have been expressed in the non-infection quantification method, 24 lipoprotein transcripts including two significantly up-regulated *SAR0706* and *SAR1402* transcripts that expressed under the infection condition did not show a detectable level of translated lipoproteins in the non-infection condition. This may reflect the real life

situation but may also be due to the different sensitivities of the two techniques and low levels of expression may not have been detected in the proteomics experiments. One of the important functions of lipoproteins is as transporters for many nutrients or chaperones, therefore it is not unexpected that the *lgt* mutation bacterium showed affected high-affinity metal ion uptake and also attenuated virulence (Schmaler *et al.*, 2009). *S. aureus* USA300 mutant in νSaα genomic island which containing tandem lipoprotein-like genes *lpl* were lacking in the stimulation of pro-inflammatory cytokines of human monocytes, macrophages and keratinocytes (Nguyen *et al.*, 2015). νSaα was not annotated on the MRSA252 chromosome from Genbank that was used here but the tandem lipoproteins SAR0438-0445 may represent a similar cluster. The potential roles of lipoproteins in infection and innate immune response was not obvious in the *C. elegans* infection experiment, ~ 4 fold increase in *SAR0706* and an almost 2 fold increase in *SAR1402* and *SAR2457* expression were the only observed lipoproteins transcripts the involved in the innate immune stimulation which comprise 5% of all lipoproteins of *S. aureus* MRSA252, however the significantly down-regulated lipoprotein genes were comprise ~18%, to suggest that these genes do not contributes in innate immune stimulation activity. Two lipoprotein genes *SAR2736* and *SAR1565* were annotated as pseudogene in the genomes and protein NCBI database, but both genes were expressed and detected in the RNA Sequencing procedure to suggest that these genes are real genes.

Most of the investigation on *S. aureus* aimed to identify particular virulence factors and novel therapeutic targets for the *S. aureus* infections, this study has focused on the important epidemic strains EMRSA-16 *S. aureus* that cause nosocomial infections in the UK and around the world. Lipoproteins are effective vaccine candidates and proved protective immunity against *S. aureus* (Mishra *et al*., 2012; Mariotti *et al*., 2013). There is high demand for new treatment in order to treat *S. aureus* infections as only few antibiotics have a proved an effectiveness to treat the MRSA strains in the last decades. Although the results presented here have demonstrated some variation of lipoproteins genes structure and expression, further work is needed to investigate the remaining lipoproteins and their expression levels. It would also be valuable to evaluate lipoprotein expression at different times of infection of *C. elegans* this would lead to a better understanding of *S. aureus* infection mechanisms and how the infections occur. It will also be necessary to determine their roles in pathogenesis in a model system that involves cell-mediated

immunity unlike *C. elegans* and find new approaches to prevent infection, develop new and more effective treatments.

7. References

Aballay, A. (2013). Role of the nervous system in the control of proteostasis during innate immune activation: insights from *C. elegans*. *PLoS Pathogens, 9***,** e1003433.

Aballay, A., Yorgey, P. & Ausubel, F. M. (2000). *Salmonella typhimurium* proliferates and establishes a persistent infection in the intestine of *Caenorhabditis elegans*. *Current Biology, 10***,** 1539-1542.

Abrahmsen, L., Moks, T., Nilsson, B., Hellman, U. & Uhlen, M. (1985). Analysis of signals for secretion in the staphylococcal protein A gene. *The EMBO Journal, 4***,** 3901.

Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature, 422***,** 198-207.

Akira, S. (2003). Mammalian Toll-like receptors. *Current Opinion In Immunology, 15***,** 5-11.

Akira, S., Uematsu, S. & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell, 124***,** 783-801.

Allard, M., Moisan, H., Brouillette, É., Gervais, A. L., Jacques, M., Lacasse, P., Diarra, M. S. & Malouin, F. (2006). Transcriptional modulation of some *Staphylococcus aureus* iron-regulated genes during growth *in vitro* and in a tissue cage model *in vivo*. *Microbes and Infection, 8***,** 1679-1690.

Alper, S., McBride, S. J., Lackford, B., Freedman, J. H. & Schwartz, D. A. (2007). Specificity and complexity of the *Caenorhabditis elegans* innate immune response. *Molecular and Cellular Biology, 27***,** 5544-5553.

Anderson, A. S., Scully, I. L., Timofeyeva, Y., Murphy, E., McNeil, L. K., Mininni, T., Nuñez, L., Carriere, M., Singer, C., Dilts, D. A. & Jansen, K. U. (2012). *Staphylococcus aureus* manganese transport protein c is a highly conserved cell surface protein that elicits protective immunity against *S. aureus* and *Staphylococcus epidermidis*. *Journal of Infectious Diseases, 205***,** 1688-1696.

Anderson, J. S., Matsuhashi, M., Haskin, M. A. & Strominger, J. L. (1965). Lipid-phosphoacetylmuramyl-pentapeptide and lipid-phosphodisaccharide-pentapeptide: presumed membrane transport intermediates in cell wall synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 53***,** 881.

Andrew, P. A. & Nicholas, W. (1976). Effect of bacteria on dispersal of *Caenorhabditis elegans* (Rhabditidae). *Nematologica, 22***,** 451-461.

Andrews, S. C., Robinson, A. K. & Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiology Reviews, 27***,** 215-237.

Appelbaum, P. C. (2007). Microbiology of antibiotic resistance in *Staphylococcus aureus*. *Clinical Infectious Diseases, 45***,** S165-S170.

Archer, G. L. (1998). *Staphylococcus aureus*: a well-armed pathogen. *Clinical Infectious Diseases, 26***,** 1179-1181.

Arvidson, S. & Tegmark, K. (2001). Regulation of virulence determinants in *Staphylococcus aureus*. *International Journal of Medical Microbiology, 291***,** 159-170.

Asanuma, M., Kurokawa, K., Ichikawa, R., Ryu, K. H., Chae, J. H., Dohmae, N., Lee, B. L. & Nakayama, H. (2011). Structural evidence of α-aminoacylated lipoproteins of *Staphylococcus aureus*. *FEBS Journal, 278***,** 716-728.

Aylin, H., Mareike, S. & Matthias, L. (2012). The saposin-like protein SPP-12 is an antimicrobial polypeptide in the pharyngeal neurons of *Caenorhabditis elegans* and participates in defence against a natural bacterial pathogen. *Biochemical Journal, 445***,** 205-212.

Baba, T., Bae, T., Schneewind, O., Takeuchi, F. & Hiramatsu, K. (2008). Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *Journal of Bacteriology, 190***,** 300-310.

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K.-i., Oguchi, A., Nagai, Y., Iwama, N., Asano, K. & Naimi, T. (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *The Lancet, 359***,** 1819-1827.

Babu, M. M., Priya, M. L., Selvan, A. T., Madera, M., Gough, J., Aravind, L. & Sankaran, K. (2006). A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *Journal of Bacteriology, 188***,** 2761-2773.

Babu, M. M. & Sankaran, K. (2002). DOLOP database of bacterial lipoproteins. *Bioinformatics, 18***,** 641-643.

Bae, T., Banger, A. K., Wallace, A., Glass, E. M., Åslund, F., Schneewind, O. & Missiakas, D. M. (2004). *Staphylococcus aureus* virulence genes identified by *bursa aurealis* mutagenesis and nematode killing. *Proceedings of the National Academy of Sciences of the United States of America, 101***,** 12312-12317.

Baggett, H. C., Hennessy, T. W., Rudolph, K., Bruden, D., Reasonover, A., Parkinson, A., Sparks, R., Donlan, R. M., Martinez, P. & Mongkolrattanothai, K. (2004). Community-onset methicillin-resistant *Staphylococcus aureus* associated with antibiotic use and the cytotoxin Panton-Valentine leukocidin during a furunculosis outbreak in rural Alaska. *Journal of Infectious Diseases, 189***,** 1565-1573.

Bagnoli, F., Fontana, M. R., Soldaini, E., Mishra, R. P., Fiaschi, L., Cartocci, E., Nardi-Dei, V., Ruggiero, P., Nosari, S. & De Falco, M. G. (2015). Vaccine composition formulated with a novel TLR7-dependent adjuvant induces high and broad protection against *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences, 112***,** 3680-3685.

Bagos, P. G., Tsirigos, K. D., Liakopoulos, T. D. & Hamodrakas, S. J. (2008). Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *Journal of Proteome Research, 7***,** 5082-5093.

Bairoch, A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research, 28***,** 45-48.

Ballal, A., Ray, B. & Manna, A. C. (2009). *sarZ*, a *sarA* family gene, is transcriptionally activated by MgrA and is involved in the regulation of genes encoding exoproteins in *Staphylococcus aureus*. *Journal of Bacteriology, 191***,** 1656-1665.

Bals, R. & Hiemstra, P. (2004). Innate immunity in the lung: how epithelial cells fight against respiratory pathogens. *European Respiratory Journal, 23***,** 327-333.

Baptiste, K. E., Williams, K., Willams, N. J., Wattret, A., Clegg, P. D., Dawson, S., Corkill, J. E., O'Neill, T. & Hart, C. A. (2005). Methicillin-resistant staphylococci in companion animals. *Emerging Infectious Diseases, 11***,** 1942.

Barber, M. (1961). Methicillin-resistant staphylococci. *Journal of Clinical Pathology, 14***,** 385.

Bartels, M., Larner-Svensson, H., Meiniche, H., Kristoffersen, K., Schonning, K., Nielsen, J., Rohde, S., Christensen, L., Skibsted, A. & Jarlov, J. (2014). Monitoring meticillin resistant Staphylococcus aureus and its spread in Copenhagen, Denmark, 2013, through routine whole genome sequencing. *Euro Surveillance: bulletin Europeen sur les maladies transmissibles= European communicable disease bulletin, 20*.

Bayer, M. G., Heinrichs, J. H. & Cheung, A. L. (1996). The molecular architecture of the *sar* locus in *Staphylococcus aureus*. *Journal of Bacteriology, 178***,** 4563-4570.

Beale, E., Li, G., Tan, M.-W. & Rumbaugh, K. P. (2006). *Caenorhabditis elegans* senses bacterial autoinducers. *Applied and Environmental Microbiology, 72***,** 5135-5137.

Beard-Pegler, M. A., Stubbs, E. & Vickery, A. M. (1988). Observations on the resistance to drying of staphylococcal strains. *Journal of Medical Microbiology, 26***,** 251-255.

Beaume, M., Hernandez, D., Docquier, M., Delucinge-Vivier, C., Descombes, P. & François, P. (2011). Orientation and expression of methicillin-resistant *Staphylococcus aureus* small RNAs by direct multiplexed measurements using the nCounter of NanoString technology. *Journal of Microbiological Methods, 84***,** 327-334.

Beaume, M., Hernandez, D., Farinelli, L., Deluen, C., Linder, P., Gaspin, C., Romby, P., Schrenzel, J. & Francois, P. (2010). Cartography of methicillin-resistant *S. aureus* transcripts: detection, orientation and temporal expression during growth phase and stress conditions. *PLoS One, 5***,** e10725.

Becher, D., Hempel, K., Sievers, S., Zühlke, D., Pané-Farré, J., Otto, A., Fuchs, S., Albrecht, D., Bernhardt, J. & Engelmann, S. (2009a). A proteomic view of an important human pathogen-towards the quantification of the entire *Staphylococcus aureus* proteome. *PLoS One, 4***,** 8176.

Becher, D., Hempel, K., Sievers, S., Zühlke, D., Pané-Farré, J., Otto, A., Fuchs, S., Albrecht, D., Bernhardt, J. & Engelmann, S. (2009b). A proteomic view of an important human pathogen–towards the quantification of the entire *Staphylococcus aureus* proteome. *PLoS One, 4***,** e8176.

Ben Nejma, M., Merghni, A. & Mastouri, M. (2014). Genotyping of Methicillin Resistant *Staphylococcus aureus* strains isolated from hospitalized children. *International Journal of Pediatrics, 2014*.

Bengtsson, J., Tjalsma, H., Rivolta, C. & Hederstedt, L. (1999). Subunit II of *Bacillus subtilis* Cytochromec Oxidase Is a Lipoprotein. *Journal of Bacteriology, 181***,** 685-688.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 289-300.

Beyer, A., Hollunder, J., Nasheuer, H.-P. & Wilhelm, T. (2004). Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Molecular and Cellular Proteomics, 3***,** 1083-1092.

Bien, J., Sokolova, O. & Bozko, P. (2011). Characterization of virulence factors of *Staphylococcus aureus*: novel function of known virulence factors that are implicated in activation of airway epithelial proinflammatory response. *Journal of Pathogens, 2011*.

Bignardi, G., Woodford, N., Chapman, A., Johnson, A. & Speller, D. (1996). Detection of the *mecA* gene and phenotypic detection of resistance in *Staphylococcus aureus* isolates with borderline or low-level methicillin resistance. *Journal of Antimicrobial Chemotherapy, 37***,** 53-63.

Bischoff, M., Entenza, J. & Giachino, P. (2001). Influence of a Functional sigB Operon on the Global Regulators sar and agr in *Staphylococcus aureus*. *Journal of Bacteriology, 183***,** 5171-5179.

Bland, J. M. & Altman, D. G. (1998). Survival probabilities (the Kaplan-Meier method). *Bmj, 317***,** 1572-1580.

Bloomfield, S. F., Cookson, B., Falkiner, F., Griffith, C. & Cleary, V. (2007). Methicillin-resistant *Staphylococcus aureus, Clostridium difficile*, and extended-spectrum β-lactamase–producing *Escherichia coli* in the community: Assessing the problem and controlling the spread. *American Journal of Infection Control, 35***,** 86-88.

Brady, R. A., Leid, J. G., Camper, A. K., Costerton, J. W. & Shirtliff, M. E. (2006). Identification of *Staphylococcus aureus* proteins recognized by the antibody-mediated immune response to a biofilm infection. *Infection and Immunity, 74***,** 3415-3426.

Braun, V. & Rehn, K. (1969). Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. *European Journal of Biochemistry, 10***,** 426-438.

Braun, V. & Wu, H. (1994). Lipoproteins, structure, function, biosynthesis and model for protein export. *New Comprehensive Biochemistry, 27***,** 319-319.

Broekema, N. M., Van, T. T., Monson, T. A., Marshall, S. A. & Warshauer, D. M. (2009). Comparison of cefoxitin and oxacillin disk diffusion methods for detection of *mecA*-mediated resistance in *Staphylococcus aureus* in a large-scale study. *Journal of Clinical Microbiology, 47***,** 217-219.

Brouillette, E., Grondin, G., Shkreta, L., Lacasse, P. & Talbot, B. G. (2003). *In vivo* and *in vitro* demonstration that *Staphylococcus aureus* is an intracellular pathogen in the presence or absence of fibronectin-binding proteins. *Microbial Pathogenesis, 35***,** 159-168.

Brown, J. S. & Holden, D. W. (2002). Iron acquisition by Gram-positive bacterial pathogens. *Microbes and Infection, 4***,** 1149-1156.

Bumann, D. (2010). Pathogen proteomes during infection: a basis for infection research and novel control strategies. *Journal of Proteomics, 73***,** 2267-2276.

Burke, K. A. & Lascelles, J. (1975). Nitrate reductase system in *Staphylococcus aureus* wild type and mutants. *Journal of Bacteriology, 123***,** 308-316.

Burton, D. C., Edwards, J. R., Horan, T. C., Jernigan, J. A. & Fridkin, S. K. (2009). Methicillin-resistant *Staphylococcus aureus* central line-associated bloodstream infections in US intensive care units, 1997-2007. *JAMA: the journal of the American Medical Association, 301***,** 727-736.

Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F. & Kjelleberg, S. (2007). Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology, 73***,** 278-288.

Cassat, J. E., Dunman, P. M., McAleese, F., Murphy, E., Projan, S. J. & Smeltzer, M. S. (2005). Comparative genomics of *Staphylococcus aureus* musculoskeletal isolates. *Journal of Bacteriology, 187***,** 576-592.

Chambers, H. F. (1997). Methicillin resistance in staphylococci: molecular and biochemical basis and clinical implications. *Clinical Microbiology Reviews, 10***,** 781-791.

Chang, Y.-F., Appel, M., Jacobson, R. H., Shin, S. J., Harpending, P., Straubinger, R., Patrican, L. A., Mohammed, H. & Summers, B. A. (1995). Recombinant OspA protects dogs against infection and disease caused by *Borrelia burgdorferi*. *Infection and Immunity, 63***,** 3543-3549.

Chatterjee, S. S. & Otto, M. (2013). Improved understanding of factors driving methicillin-resistant *Staphylococcus aureus* epidemic waves. *Clinical Epidemiology, 5***,** 205.

Chávez, V., Mohri-Shiomi, A. & Garsin, D. A. (2009). Ce-Duox1/BLI-3 generates reactive oxygen species as a protective innate immune mechanism in *Caenorhabditis elegans*. *Infection and Immunity, 77***,** 4983-4989.

Chávez, V., Mohri-Shiomi, A., Maadani, A., Vega, L. A. & Garsin, D. A. (2007). Oxidative stress enzymes are required for DAF-16-mediated immunity due to generation of reactive oxygen species by *Caenorhabditis elegans*. *Genetics, 176***,** 1567-1577.

Chelius, D. & Bondarenko, P. V. (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research, 1***,** 317-323.

Chen, C. J. & Huang, Y. C. (2014). New epidemiology of *Staphylococcus aureus* infection in Asia. *Clinical Microbiology and Infection, 20***,** 605-623.

Cheung, Eberhardt, K. & Heinrichs, J. H. (1997). Regulation of protein A synthesis by the sar and agr loci of *Staphylococcus aureus*. *Infection and Immunity, 65***,** 2243-2249.

Cheung, A. L., Bayer, A. S., Zhang, G., Gresham, H. & Xiong, Y. Q. (2004). Regulation of virulence determinants *in vitro* and *in vivo* in *Staphylococcus aureus*. *FEMS Immunology & Medical Microbiology, 40***,** 1-9.

Cheung, A. L., Koomey, J. M., Butler, C. A., Projan, S. J. & Fischetti, V. A. (1992). Regulation of exoprotein expression in *Staphylococcus aureus* by a locus (*sar*) distinct from *agr*. *Proceedings of the National Academy of Sciences, 89***,** 6462-6466.

Cheung, A. L. & Projan, S. J. (1994). Cloning and sequencing of sarA of *Staphylococcus aureus*, a gene required for the expression of agr. *Journal of Bacteriology, 176***,** 4168-4172.

Cheung, A. L., Schmidt, K., Bateman, B. & Manna, A. C. (2001). SarS, a SarA homolog repressible by agr, is an activator of protein A synthesis in *Staphylococcus aureus*. *Infection and Immunity, 69***,** 2448-2455.

Cheung., Bayer, M. G. & Heinrichs, J. H. (1997). *sar* genetic determinants necessary for transcription of RNAII and RNAIII in the *agr* locus of *Staphylococcus aureus*. *Journal of Bacteriology, 179***,** 3963-3971.

Chien, Y.-t. & Cheung, A. L. (1998). Molecular interactions between two global regulators, *sar* and *agr*, in *Staphylococcus aureus*. *Journal of Biological Chemistry, 273***,** 2645-2652.

Chien, Y.-t., Manna, A. C., Projan, S. J. & Cheung, A. L. (1999). *SarA*, a global regulator of virulence determinants in *Staphylococcus aureus*, binds to a conserved motif essential for *sar*-dependent gene regulation. *Journal of Biological Chemistry, 274***,** 37169-37176.

Chiorazzi, M., Rui, L., Yang, Y., Ceribelli, M., Tishbi, N., Maurer, C. W., Ranuncolo, S. M., Zhao, H., Xu, W. & Chan, W.-C. C. (2013). Related F-box proteins control cell death in *Caenorhabditis elegans* and human lymphoma. *Proceedings of the National Academy of Sciences, 110***,** 3943-3948.

Clarke, S. R., Brummell, K. J., Horsburgh, M. J., McDowell, P. W., Mohamad, S. A. S., Stapleton, M. R., Acevedo, J., Read, R. C., Day, N. P. & Peacock, S. J. (2006). Identification of *in vivo*-expressed antigens of *Staphylococcus aureus* and their use in vaccinations for protection against nasal carriage. *Journal of Infectious Diseases, 193***,** 1098-1108.

Clarke, S. R. & Foster, S. J. (2006). Surface adhesins of *Staphylococcus aureus*. *Advances in Microbial Physiology, 51***,** 187-224.

Claverys, J.-P. (2001). A new family of high-affinity ABC manganese and zinc permeases. *Research in Microbiology, 152***,** 231-243.

Clements, M. O., Watson, S. P. & Foster, S. J. (1999). Characterization of the major superoxide dismutase of *Staphylococcus aureus* and its role in starvation survival, stress resistance, and pathogenicity. *Journal of Bacteriology, 181***,** 3898-3903.

Cloonan, N. & Grimmond, S. M. (2008). Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol, 9***,** 234.

Cockayne, A., Hill, P. J., Powell, N. B., Bishop, K., Sims, C. & Williams, P. (1998). Molecular cloning of a 32-kilodalton lipoprotein component of a novel iron-regulated Staphylococcus epidermidis ABC transporter. *Infection and Immunity, 66***,** 3767-3774.

Consortium, U. (2014). UniProt: a hub for protein information. *Nucleic Acids Research***,** gku989.

Control, C. f. D. & Prevention (1997). Interim guidelines for prevention and control of Staphylococcal infection associated with reduced susceptibility to vancomycin. *MMWR. Morbidity and mortality weekly report, 46***,** 626.

Conway, T. (2003). Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Molecular Microbiology, 47***,** 879-889.

Cooper, B., Stone, S., Kibbler, C., Cookson, B., Roberts, J., Medley, G., Duckworth, G., Lai, R. & Ebrahim, S. (2004). Isolation measures in the hospital management of

methicillin resistant *Staphylococcus aureus* (MRSA): systematic review of the literature. *Bmj, 329***,** 533.

Cooper, J. E. & Feil, E. J. (2006). The phylogeny of *Staphylococcus aureus*–which genes make the best intra-species markers? *Microbiology, 152***,** 1297-1305.

Cordwell, S. J., Larsen, M. R., Cole, R. T. & Walsh, B. J. (2002). Comparative proteomics of *Staphylococcus aureus* and the response of methicillin-resistant and methicillin-sensitive strains to Triton X-100. *Microbiology, 148***,** 2765-2781.

Costa, M. O. C., Beltrame, C. O., Ferreira, F. A., Botelho, A. M. N., Lima, N. C. B., Souza, R. C., de Almeida, L. G. P., Vasconcelos, A. T. R., Nicolás, M. F. & Figueiredo, A. M. S. (2013). Complete genome sequence of a variant of the methicillin-resistant *Staphylococcus aureus* ST239 lineage, strain BMB9393, displaying superior ability to accumulate ica-independent biofilm. *Genome Announcements, 1***,** e00576-13.

Couillault, C., Pujol, N., Reboul, J., Sabatier, L., Guichou, J.-F., Kohara, Y. & Ewbank, J. J. (2004). TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. *Nature Immunology, 5***,** 488-494.

Cox, J. & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology, 26***,** 1367-1372.

Cravatt, B. F., Simon, G. M. & Yates Iii, J. R. (2007). The biological impact of mass-spectrometry-based proteomics. *Nature, 450***,** 991-1000.

Croucher, N. J. & Thomson, N. R. (2010). Studying bacterial transcriptomes using RNA-seq. *Current Opinion in Microbiology, 13***,** 619-624.

Dahl, C. E., Sacktor, N. & Dahl, J. S. (1985). Acylated proteins in *Acholeplasma laidlawii*. *Journal of Bacteriology, 162***,** 445-447.

Dalbey, R. E., Lively, M. O., Bron, S. & Dijl, J. M. V. (1997). The chemistry and enzymology of the type I signal peptidases. *Protein Science, 6***,** 1129-1138.

Darby, C., Hsu, J. W., Ghori, N. & Falkow, S. (2002). *Caenorhabditis elegans*: plague bacteria biofilm blocks food intake. *Nature, 417***,** 243-244.

David, M. Z. & Daum, R. S. (2010). Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. *Clinical Microbiology Reviews, 23***,** 616-687.

David, M. Z., Siegel, J. D., Henderson, J., Leos, G., Lo, K., Iwuora, J., Taylor, A. R., Zychowski, D. L., Porsa, E. & Boyle-Vavra, S. (2014). Hand and Nasal Carriage of Discordant *Staphylococcus aureus* Isolates among Urban Jail Detainees. *Journal Of Clinical Microbiology, 52***,** 3422-3425.

David, M. Z., Taylor, A., Lynfield, R., Boxrud, D. J., Short, G., Zychowski, D., Boyle-Vavra, S. & Daum, R. S. (2013). Comparing pulsed-field gel electrophoresis with multilocus sequence typing, spa typing, staphylococcal cassette chromosome *mec* (SCC*mec*) typing, and PCR for Panton-Valentine leukocidin, *arcA*, and *opp3* in methicillin-resistant *Staphylococcus aureus* isolates at a US medical center. *Journal of Clinical Microbiology, 51***,** 814-819.

DeLeo, F. R., Diep, B. A. & Otto, M. (2009). Host defense and pathogenesis in *Staphylococcus aureus* infections. *Infectious Disease Clinics of North America, 23***,** 17-34.

Dengler, V., Meier, P. S., Heusser, R., Berger-Bächi, B. & McCallum, N. (2011). Induction kinetics of the *Staphylococcus aureus* cell wall stress stimulon in response to different cell wall active antibiotics. *BMC Microbiology, 11***,** 16.

Desai, R., Pannaraj, P. S., Agopian, J., Sugar, C. A., Liu, G. Y. & Miller, L. G. (2011). Survival and transmission of community-associated methicillin-resistant *Staphylococcus aureus* from fomites. *American Journal of Infection Control, 39***,** 219-225.

Desvaux, M., Hébraud, M., Talon, R. & Henderson, I. R. (2009). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends in Microbiology, 17***,** 139-145.

DeVries, A. S., Lesher, L., Schlievert, P. M., Rogers, T., Villaume, L. G., Danila, R. & Lynfield, R. (2011). Staphylococcal toxic shock syndrome 2000–2006: epidemiology, clinical features, and molecular characteristics. *PLoS One, 6***,** e22997.

Diep, B. A., Phung, Q., Date, S., Arnott, D., Bakalarski, C., Xu, M., Nakamura, G., Swem, D. L., Alexander, M. K. & Le, H. N. (2014a). Identifying potential therapeutic targets of methicillin-resistant *Staphylococcus aureus* through *in vivo* proteomic analysis. *Journal of Infectious Diseases, 209***,** 1533-1541.

Diep, B. A., Phung, Q., Date, S., Arnott, D., Bakalarski, C., Xu, M., Nakamura, G., Swem, D. L., Alexander, M. K., Le, H. N., Mai, T. T., Tan, M. W., Brown, E. J. & Nishiyama, M. (2014b). Identifying potential therapeutic targets of methicillin-resistant *staphylococcus aureus* through *in vivo* proteomic analysis. *Journal of Infectious Diseases, 209***,** 1533-1541.

Dilks, K., Rose, R. W., Hartmann, E. & Pohlschröder, M. (2003). Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. *Journal of Bacteriology, 185***,** 1478-1483.

Dinges, M. M., Orwin, P. M. & Schlievert, P. M. (2000). Exotoxins of *Staphylococcus aureus*. *Clinical Microbiology Reviews, 13***,** 16-34.

Driessen, A. J. M. & Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. *Annu. Rev. Biochem., 77***,** 643-667.

Drummelsmith, J., Winstall, E., Bergeron, M. G., Poirier, G. G. & Ouellette, M. (2007). Comparative proteomics analyses reveal a potential biomarker for the detection of vancomycin-intermediate *Staphylococcus aureus* strains. *Journal of Proteome Research, 6***,** 4690-4702.

Dryden, M. S. (2014). Novel antibiotic treatment for skin and soft tissue infection. *Current Opinion In Infectious Diseases, 27***,** 116-124.

Duerden, B., Fry, C., Johnson, A. P. & Wilcox, M. H. The Control of Methicillin-Resistant *Staphylococcus aureus* Blood Stream Infections in England. Open Forum Infectious Diseases, 2015. Oxford University Press, ofv035.

Dunman, P. á., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., Brown, E., Zagursky, R., Shlaes, D. & Projan, S. (2001). Transcription Profiling-Based

Identification of *Staphylococcus aureus* Genes Regulated by the agrand/or *sarA* Loci. *Journal of Bacteriology, 183***,** 7341-7353.

Edgar, L. G., Wolf, N. & Wood, W. B. (1994). Early transcription in *Caenorhabditis elegans* embryos. *Development, 120***,** 443-451.

Edwards-Jones, V. & Foster, H. A. (2002). Effects of silver sulphadiazine on the production of exoproteins by *Staphylococcus aureus*. *Journal of Medical Microbiology, 51***,** 50-55.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science***,** 54-75.

Engelmann, I. & Pujol, N. 2010. Innate immunity in *C. elegans*. *Invertebrate Immunity.* Springer.

Evenson, M. L., Hinds, M. W., Bernstein, R. S. & Bergdoll, M. S. (1988). Estimation of human dose of staphylococcal enterotoxin A from a large outbreak of staphylococcal food poisoning involving chocolate milk. *International Journal of Food Microbiology, 7***,** 311-316.

Ewbank, J. J. & Zugasti, O. (2011). *C. elegans*: model host and tool for antimicrobial drug discovery. *Disease Models & Mechanisms, 4***,** 300-304.

Fedtke, I., Götz, F. & Peschel, A. (2004). Bacterial evasion of innate host defenses-the *Staphylococcus aureus* lesson. *International Journal of Medical Microbiology, 294***,** 189-194.

Félix, M.-A. & Braendle, C. (2010). The natural history of *Caenorhabditis elegans*. *Current Biology, 20***,** R965-R969.

Felsenstein, J. (1997). An alternating least squares approach to inferring phylogenies from pairwise distances. *Systematic Biology, 46***,** 101-111.

Feltcher, M. E., Gibbons, H. S., Ligon, L. S. & Braunstein, M. (2013). Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. *Journal of Bacteriology, 195***,** 672-681.

FigTree (2014). Available from: http://tree.bio.ed.ac.uk/software/figtree/.

Filatov, D. A. (2002). ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. *Molecular Ecology Notes, 2***,** 621-624.

Fischer, E. A. (1988). Simultaneous hermaphroditism, tit-for-tat, and the evolutionary stability of social systems. *Ethology and Sociobiology, 9***,** 119-136.

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. (2001). Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proceedings of the National Academy of Sciences, 98***,** 8821-8826.

Foster, T. J. (2005). Immune evasion by staphylococci. *Nature Reviews Microbiology, 3***,** 948-958.

Foster, T. J. & Höök, M. (1998). Surface protein adhesins of *Staphylococcus aureus*. *Trends in Microbiology, 6***,** 484-488.

Fournier, B. & Hooper, D. C. (2000). A New Two-Component Regulatory System Involved in Adhesion, Autolysis, and Extracellular Proteolytic Activity of *Staphylococcus aureus*. *Journal of Bacteriology, 182***,** 3955-3964.

Fournier, B. & Philpott, D. J. (2005). Recognition of *Staphylococcus aureus* by the innate immune system. *Clinical Microbiology Reviews, 18***,** 521-540.

Frénay, H., Bunschoten, A., Schouls, L., Van Leeuwen, W., Vandenbroucke-Grauls, C., Verhoef, J. & Mooi, F. (1996). Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *European Journal of Clinical Microbiology and Infectious Diseases, 15***,** 60-64.

Fröderberg, L., Houben, E. N., Baars, L., Luirink, J. & De Gier, J.-W. (2004). Targeting and translocation of two lipoproteins in *Escherichia coli* via the SRP/Sec/YidC pathway. *Journal of Biological Chemistry, 279***,** 31026-31032.

Fuchs, S., Pané-Farré, J., Kohler, C., Hecker, M. & Engelmann, S. (2007). Anaerobic gene expression in *Staphylococcus aureus*. *Journal of Bacteriology, 189***,** 4275-4289.

Fueyo, J., Mendoza, M. C., Rodicio, M. R., Muniz, J., Alvarez, M. & Martín, M. C. (2005). Cytotoxin and pyrogenic toxin superantigen gene profiles of *Staphylococcus aureus* associated with subclinical mastitis in dairy cows and relationships with macrorestriction genomic profiles. *Journal of Clinical Microbiology, 43***,** 1278-1284.

Gallagher, L. A. & Manoil, C. (2001). *Pseudomonas aeruginosa* PAO1 kills *Caenorhabditis elegans* by cyanide poisoning. *Journal of Bacteriology, 183***,** 6207-6214.

García-Álvarez, L., Holden, M. T., Lindsay, H., Webb, C. R., Brown, D. F., Curran, M. D., Walpole, E., Brooks, K., Pickard, D. J. & Teale, C. (2011). Meticillin-resistant *Staphylococcus aureus* with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. *The Lancet Infectious Diseases, 11***,** 595-603.

Gardner, M., Rosell, M. & Myers, E. M. (2013). Measuring the Effects of Bacteria on *C. Elegans* Behavior Using an Egg Retention Assay. *JoVE (Journal of Visualized Experiments)***,** e51203-e51203.

Garsin, D. A., Sifri, C. D., Mylonakis, E., Qin, X., Singh, K. V., Murray, B. E., Calderwood, S. B. & Ausubel, F. M. (2001). A simple model host for identifying Gram-positive virulence factors. *Proceedings of the National Academy of Sciences, 98***,** 10892-10897.

Gatlin, C. L., Pieper, R., Huang, S. T., Mongodin, E., Gebregeorgis, E., Parmar, P. P., Clark, D. J., Alami, H., Papazisi, L. & Fleischmann, R. D. (2006). Proteomic profiling of cell envelope-associated proteins from *Staphylococcus aureus*. *Proteomics, 6***,** 1530-1549.

Georgopapadakou, N. H. & Liu, F. Y. (1980). Binding of beta-lactam antibiotics to penicillin-binding proteins of *Staphylococcus aureus* and Streptococcus faecalis: relation to antibacterial activity. *Antimicrobial Agents and Chemotherapy, 18***,** 834-836.

Ghuysen, J.-M. (1994). Molecular structures of penicillin-binding proteins and β-lactamases. *Trends in Microbiology, 2***,** 372-380.

Ghuysen, J.-M. & Strominger, J. L. (1963). Structure of the cell wall of *Staphylococcus aureus*, strain Copenhagen. I. Preparation of fragments by enzymatic hydrolysis. *Biochemistry, 2***,** 1110-1119.

Gilbert, J. A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P. & Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PloS One, 3***,** e3042.

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L. & Beanan, M. (2005). Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *Journal of Bacteriology, 187***,** 2426-2438.

Gillet, Y., Issartel, B., Vanhems, P., Fournet, J.-C., Lina, G., Bes, M., Vandenesch, F., Piémont, Y., Brousse, N. & Floret, D. (2002). Association between *Staphylococcus aureus* strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *The Lancet, 359***,** 753-759.

Giombini, E., Orsini, M., Carrabino, D. & Tramontano, A. (2010). An automatic method for identifying surface proteins in bacteria: SLEP. *BMC Bioinformatics, 11***,** 39.

Giraudo, A. T., Raspanti, C. G., Calzolari, A. & Nagel, R. (1994). Characterization of a Tn 551-mutant of *Staphylococcus aureus* defective in the production of several exoproteins. *Canadian Journal of Microbiology, 40***,** 677-681.

Golden, T. R. & Melov, S. (2007). Gene expression changes associated with aging in *C. elegans*.

Gordon, C. P., Williams, P. & Chan, W. C. (2013). Attenuating *Staphylococcus aureus* virulence gene regulation: a medicinal chemistry perspective. *Journal of Medicinal Chemistry, 56***,** 1389-1404.

Gordon, N., Price, J., Cole, K., Everitt, R., Morgan, M., Finney, J., Kearns, A., Pichon, B., Young, B. & Wilson, D. (2014). Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *Journal of Clinical Microbiology, 52***,** 1182-1191.

Gorg, A., Weiss, W. & Dunn, M. J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics, 4***,** 3665-3685.

Gouy, M., Guindon, S. & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution, 27***,** 221-224.

Grandi, G. (2010). Bacterial surface proteins and vaccines. *F1000 Biology Reports, 2*.

Gravato-Nobre, M. J. & Hodgkin, J. (2005). *Caenorhabditis elegans* as a model for innate immunity to pathogens. *Cellular Microbiology, 7***,** 741-751.

Greenbaum, D., Jansen, R. & Gerstein, M. (2002). Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics, 18***,** 585-596.

Grundmeier, M., Tuchscherr, L., Brück, M., Viemann, D., Roth, J., Willscher, E., Becker, K., Peters, G. & Löffler, B. (2010). Staphylococcal strains vary greatly in their ability to induce an inflammatory response in endothelial cells. *Journal of Infectious Diseases, 201***,** 871-880.

Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T. & Kühner, S. (2009). Transcriptome complexity in a genome-reduced bacterium. *Science, 326***,** 1268-1271.

Güell, M., Yus, E., Lluch-Senar, M. & Serrano, L. (2011). Bacterial transcriptomics: what is beyond the RNA horiz-ome? *Nature Reviews Microbiology, 9***,** 658-669.

Hajjeh, R. A., Reingold, A., Weil, A., Shutt, K., Schuchat, A. & Perkins, B. A. (1999). Toxic shock syndrome in the United States: surveillance update, 1979 1996. *Emerging Infectious Diseases, 5***,** 807.

Hanson, M. S., Cassatt, D. R., Guo, B. P., Patel, N. K., McCarthy, M. P., Dorward, D. W. & Höök, M. (1998). Active and passive immunity against *Borrelia burgdorferi* decorin binding protein A (DbpA) protects against infection. *Infection and Immunity, 66***,** 2143-2153.

Hanssen, A. M. & Ericson Sollid, J. U. (2006). SCC*mec* in staphylococci: genes on the move. *FEMS Immunology & Medical Microbiology, 46***,** 8-20.

Hantke, K. (2001). Iron and metal regulation in bacteria. *Current Opinion in Microbiology, 4***,** 172-177.

Hantke, K. & Braun, V. (1973). Covalent binding of lipid to protein. *European Journal of Biochemistry, 34***,** 284-296.

Harris, L., Foster, S. & Richards, R. (2002). An introduction to *Staphylococcus aureus*, and techniques for identifying and quantifying *S. aureus* adhesins in relation to adhesion to biomaterials: review. *Eur Cell Mater, 4***,** 39-60.

Hartman, B. J. & Tomasz, A. (1984). Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. *Journal of Bacteriology, 158***,** 513-516.

Hashimoto, M., Tawaratsumida, K., Kariya, H., Aoyama, K., Tamura, T. & Suda, Y. (2006a). Lipoprotein is a predominant Toll-like receptor 2 ligand in *Staphylococcus aureus* cell wall components. *International Immunology, 18***,** 355-362.

Hashimoto, M., Tawaratsumida, K., Kariya, H., Kiyohara, A., Suda, Y., Krikae, F., Kirikae, T. & Götz, F. (2006b). Not lipoteichoic acid but lipoproteins appear to be the dominant immunobiologically active compounds in *Staphylococcus aureus*. *The Journal of Immunology, 177***,** 3162-3169.

Hayashi, S. & Wu, H. C. (1990). Lipoproteins in bacteria. *Journal of bioenergetics and biomembranes, 22***,** 451-471.

Health Protection Agency 2009 Meticillin Resistant *Staphylococcus aureus* (MRSA) Screening and Suppression. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/330793/MRSA_screening_and_supression_primary_care_guidance.pdf. *Quick Reference Guide for Primary Care for consultation and local adaptation.*

Health Protection Agency 2015. Mandatory surveillance of *Staphylococcus aureus* bacteraemia. . *Quick Reference Guide for Primary Care for consultation and local adaptation.* http://www.hpa.org.uk/web/hpaweb&page&hpawebautolistname/page/1191942169773. Accessed October 21, 2015.

Hecker, M., Becher, D., Fuchs, S. & Engelmann, S. (2010). A proteomic view of cell physiology and virulence of *Staphylococcus aureus*. *International Journal of Medical Microbiology, 300***,** 76-87.

Heijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *European Journal of Biochemistry, 133***,** 17-21.

Heikkinen, O., Seppala, R., Tossavainen, H., Heikkinen, S., Koskela, H., Permi, P. & Kilpeläinen, I. (2009). Solution structure of the parvulin-type PPIase domain of *Staphylococcus aureus* PrsA–implications for the catalytic mechanism of parvulins. *BMC Structural Biology, 9***,** 17.

Hempel, K., Herbst, F. A., Moche, M., Hecker, M. & Becher, D. (2011). Quantitative proteomic view on secreted, cell surface-associated, and cytoplasmic proteins of the methicillin-resistant human pathogen *staphylococcus aureus* under iron-limited conditions. *Journal of Proteome Research, 10***,** 1657-1666.

Herron-Olson, L., Fitzgerald, J. R., Musser, J. M. & Kapur, V. (2007). Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS One, 2***,** e1120.

Hertz, G. Z. & Stormo, G. D. (1995). *Escherichia coli* promoter sequences: analysis and prediction. *Methods in Enzymology, 273***,** 30-42.

Herzer, C. M. (2001). Toxic shock syndrome: broadening the differential diagnosis. *The Journal of the American Board of Family Practice, 14***,** 131-136.

Hess, P. N. & De Moraes Russo, C. A. (2007). An empirical test of the midpoint rooting method. *Biological Journal of the Linnean Society, 92***,** 669-674.

Hierholzer, W., Garner, J. S., Adams, A. B., Craven, D., Fleming, D., Forlenza, S., Gilchrist, M., Goldmann, D., Larson, E. & Mayhall, C. (1995). Recommendations for preventing the spread of vancomycin resistance: recommendations of the Hospital Infection Control Practices Advisory Committee (HICPAC). *Am J Infect Control, 23***,** 87-94.

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T. & Tenover, F. (1997). Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. *Journal of Antimicrobial Chemotherapy, 40***,** 135-136.

Hiramatsu, K., Watanabe, S., Takeuchi, F., Ito, T. & Baba, T. (2004). Genetic characterization of methicillin-resistant *Staphylococcus aureus*. *Vaccine, 22***,** S5-S8.

Hodgkin, J. (1986). Sex determination in the nematode *C. elegans*: analysis of tra-3 suppressors and characterization of fem genes. *Genetics, 114***,** 15-52.

Hodgkin, J., Kuwabara, P. E. & Corneliussen, B. (2000). A novel bacterial pathogen, *Microbacterium nematophilum*, induces morphological change in the nematode *C. elegans*. *Current Biology, 10***,** 1615-1618.

Holden, M. T., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L. & Atkin, R. (2004). Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America, 101***,** 9786-9791.

Holden, M. T., Lindsay, J. A., Corton, C., Quail, M. A., Cockfield, J. D., Pathak, S., Batra, R., Parkhill, J., Bentley, S. D. & Edgeworth, J. D. (2010). Genome sequence of a

recently emerged, highly transmissible, multi-antibiotic-and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). *Journal of Bacteriology, 192***,** 888-892.

Holtfreter, S. & Broker, B. (2005). Staphylococcal superantigens: do they play a role in sepsis. *Arch Immunol Ther Exp (Warsz), 53***,** 13-27.

Höltje, J.-V. (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiology and Molecular Biology Reviews, 62***,** 181-203.

Horsburgh, M. J., Wharton, S. J., Cox, A. G., Ingham, E., Peacock, S. & Foster, S. J. (2002). MntR modulates expression of the PerR regulon and superoxide resistance in *Staphylococcus aureus* through control of manganese uptake. *Molecular Microbiology, 44***,** 1269-1286.

Howe, R. & Andrews, J. (2012). BSAC standardized disc susceptibility testing method (version 11). *Journal of Antimicrobial Chemotherapy, 67***,** 2783-2784.

Huang, T. W., Chen, F. J., Miu, W. C., Liao, T. L., Lin, A. C., Huang, I. W., Wu, K. M., Tsai, S. F., Chen, Y. T. & Lauderdale, T. L. Y. (2012). Complete genome sequence of *Staphylococcus aureus* M013, a *pvl*-positive, ST59-SCC*mec* type V strain isolated in Taiwan. *Journal of Bacteriology, 194***,** 1256-1257.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B. A., De Castro, E., Lachaize, C., Langendijk-Genevaux, P. S. & Sigrist, C. J. (2008). The 20 years of PROSITE. *Nucleic Acids Research, 36***,** D245-D249.

Hustoft, H. K., Malerod, H., Wilson, S. R., Reubsaet, L., Lundanes, E. & Greibrokk, T. (2012). A critical review of trypsin digestion for LC-MS based proteomics. *Integrative Proteomics***,** 73.

Hutchings, M. I., Palmer, T., Harrington, D. J. & Sutcliffe, I. C. (2009). Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends in Microbiology, 17***,** 13-21.

Iandolo, J. J., Worrell, V., Groicher, K. H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S. & Loh, P. (2002). Comparative analysis of the genomes of the temperate bacteriophages φ11, φ12 and φ13 of *Staphylococcus aureus* 8325. *Gene, 289***,** 109-118.

Ibarra, J. A., Pérez-Rueda, E., Carroll, R. K. & Shaw, L. N. (2013). Global analysis of transcriptional regulators in *Staphylococcus aureus*. *BMC Genomics, 14***,** 126.

Inouye, S., Wang, S. S., Sekizawa, J., Halegoua, S. & Inouye, M. (1977). Amino acid sequence for the peptide extension on the prolipoprotein of the *Escherichia coli* outer membrane. *Proceedings of the National Academy of Sciences, 74***,** 1004-1008.

Irazoqui, J. E., Troemel, E. R., Feinbaum, R. L., Luhachack, L. G., Cezairliyan, B. O. & Ausubel, F. M. (2010a). Distinct pathogenesis and host responses during infection of *C. elegans* by *P. aeruginosa* and *S. aureus*. *PLoS Pathogens, 6***,** e1000982.

Irazoqui, J. E., Urbach, J. M. & Ausubel, F. M. (2010b). Evolution of host innate defence: insights from *Caenorhabditis elegans* and primitive invertebrates. *Nature Reviews Immunology, 10***,** 47-58.

Iwasaki, A. & Medzhitov, R. (2004). Toll-like receptor control of the adaptive immune responses. *Nature Immunology, 5***,** 987-995.

Iwasaki, A. & Medzhitov, R. (2010). Regulation of adaptive immunity by the innate immune system. *Science, 327***,** 291-295.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3***,** 318-356.

Jakubovics, N. S. & Jenkinson, H. F. (2001). Out of the iron age: new insights into the critical role of manganese homeostasis in bacteria. *Microbiology, 147***,** 1709-1718.

Janini, G. M. & Veenstra, T. D. (2002). Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis, 23***,** 3048-3061.

Jansen, G., Weinkove, D. & Plasterk, R. H. (2002). The G-protein γ subunit *gpc-1* of the nematode *C. elegans* is involved in taste adaptation. *The EMBO journal, 21***,** 986-994.

Jansson, H.-B. (1994). Adhesion of conidia of *Drechmeria coniospora* to *Caenorhabditis elegans* wild type and mutants. *Journal of Nematology, 26***,** 430.

Jansson, H.-B., Jeyaprakash, A. & Zuckerman, B. M. (1985). Differential adhesion and infection of nematodes by the endoparasitic fungus *Meria coniospora* (*Deuteromycetes*). *Applied and Environmental Microbiology, 49***,** 552-555.

Jarvis, W. R., Jarvis, A. A. & Chinn, R. Y. (2012). National prevalence of methicillin-resistant *Staphylococcus aureus* in inpatients at United States health care facilities, 2010. *American Journal of Infection Control, 40***,** 194-200.

JebaMercy, G. & Balamurugan, K. (2012). Effects of sequential infections of *Caenorhabditis elegans* with *Staphylococcus aureus* and *Proteus mirabilis*. *Microbiology and Immunology, 56***,** 825-835.

JebaMercy, G., Pandian, S. K. & Balamurugan, K. (2011). Changes in *Caenorhabditis elegans* life span and selective innate immune genes during *Staphylococcus aureus* infection. *Folia Microbiologica, 56***,** 373-380.

Jia, K., Thomas, C., Akbar, M., Sun, Q., Adams-Huet, B., Gilpin, C. & Levine, B. (2009). Autophagy genes protect against *Salmonella typhimurium* infection and mediate insulin signaling-regulated pathogen resistance. *Proceedings of the National Academy of Sciences, 106***,** 14564-14569.

Johnson, A. P., Aucken, H. M., Cavendish, S., Ganner, M., Wale, M. C., Warner, M., Livermore, D. M. & Cookson, B. D. (2001). Dominance of EMRSA-15 and-16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS). *Journal of Antimicrobial Chemotherapy, 48***,** 143-144.

Jonsson, M., Mazmanian, S. K., Schneewind, O., Verdrengh, M., Bremell, T. & Tarkowski, A. (2002). On the role of *Staphylococcus aureus* sortase and sortase-catalyzed surface protein anchoring in murine septic arthritis. *Journal of Infectious Diseases, 185***,** 1417-1424.

Jonsson, P., Lindberg, M., Haraldsson, I. & Wadström, T. (1985). Virulence of *Staphylococcus aureus* in a mouse mastitis model: studies of alpha hemolysin, coagulase, and protein A as possible virulence determinants with protoplast fusion and gene cloning. *Infection and Immunity, 49***,** 765-769.

Joshua, G., Karlyshev, A., Smith, M., Isherwood, K., Titball, R. & Wren, B. (2003). A *Caenorhabditis elegans* model of *Yersinia* infection: biofilm formation on a biotic surface. *Microbiology, 149***,** 3221-3229.

Jousselin, A., Renzoni, A., Andrey, D. O., Monod, A., Lew, D. P. & Kelley, W. L. (2012). The posttranslocational chaperone lipoprotein PrsA is involved in both glycopeptide and oxacillin resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy, 56***,** 3629-3640.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2009). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Science, 12***,** 1652-1662.

Kallen, A. J., Mu, Y., Bulens, S., Reingold, A., Petit, S., Gershman, K., Ray, S. M., Harrison, L. H., Lynfield, R. & Dumyati, G. (2010). Health care-associated invasive MRSA infections, 2005-2008. *JAMA: the journal of the American Medical Association, 304***,** 641-647.

Kamath, R. S. & Ahringer, J. (2003). Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods, 30***,** 313-321.

Kang, H. J., Ha, J.-m., Kim, H. S., Lee, H., Kurokawa, K. & Lee, B. L. (2011). The role of phagocytosis in IL-8 production by human monocytes in response to lipoproteins on *Staphylococcus aureus*. *Biochemical and Biophysical Research Communications, 406***,** 449-453.

Kanoksil, M., Jatapai, A., Peacock, S. J. & Limmathurotsakul, D. (2013). Epidemiology, microbiology and mortality associated with community-acquired bacteremia in northeast Thailand: a multicenter surveillance study. *PLoS One, 8***,** e54714.

Katayama, Y., Ito, T. & Hiramatsu, K. (2000). A new class of genetic element, staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy, 44***,** 1549-1555.

Keenan, J., Oliaro, J., Domigan, N., Potter, H., Aitken, G., Allardyce, R. & Roake, J. (2000). Immune Response to an 18-Kilodalton Outer Membrane Antigen Identifies Lipoprotein 20 as a *Helicobacter pyloriVaccine* Candidate. *Infection and Immunity, 68***,** 3337-3343.

Khokhlova, O. E., Hung, W.-C., Wan, T.-W., Iwao, Y., Takano, T., Higuchi, W., Yachenko, S. V., Teplyakova, O. V., Kamshilova, V. V. & Kotlovsky, Y. V. (2015). Healthcare-and Community-Associated Methicillin-Resistant *Staphylococcus aureus* (MRSA) and Fatal Pneumonia with Pediatric Deaths in Krasnoyarsk, Siberian Russia: Unique MRSA's Multiple Virulence Factors, Genome, and Stepwise Evolution. *PloS One, 10*.

Kim, B.-S., Yi, H., Chun, J. & Cha, C.-J. (2014). Genome sequence of type strain of *Staphylococcus aureus* subsp. . *Gut Pathogens, 6***,** 1-5.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S. L. (2013a). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol, 14***,** R36.

Kim, J., Yang, J., Park, O. J., Kang, S. S., Kim, W. S., Kurokawa, K., Yun, C. H., Kim, H. H., Lee, B. L. & Han, S. H. (2013b). Lipoproteins are an important bacterial component responsible for bone destruction through the induction of osteoclast

differentiation and activation. *Journal of Bone and Mineral Research, 28***,** 2381-2391.

Kim, N. J., Ahn, K. B., Jeon, J. H., Yun, C. H., Finlay, B. B. & Han, S. H. (2015). Lipoprotein in the cell wall of *Staphylococcus aureus* is a major inducer of nitric oxide production in murine macrophages. *Molecular Immunology, 65***,** 17-24.

Klevens, R. M., Morrison, M. A., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L. H., Lynfield, R., Dumyati, G. & Townes, J. M. (2007). Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *Jama, 298***,** 1763-1771.

Kloos, W. E. (1980). Natural populations of the genus Staphylococcus. *Annual Reviews in Microbiology, 34***,** 559-592.

Kluytmans, J., Van Belkum, A. & Verbrugh, H. (1997). Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clinical Microbiology Reviews, 10***,** 505-520.

Knox, K. & Wicken, A. (1973). Immunological properties of teichoic acids. *Bacteriological Reviews, 37***,** 215.

Kohler, C., von Eiff, C., Peters, G., Proctor, R. A., Hecker, M. & Engelmann, S. (2003). Physiological characterization of a heme-deficient mutant of *Staphylococcus aureus* by a proteomic approach. *Journal of Bacteriology, 185***,** 6928-6937.

Kohler, C., Wolff, S., Albrecht, D., Fuchs, S., Becher, D., Büttner, K., Engelmann, S. & Hecker, M. (2005). Proteome analyses of *Staphylococcus aureus* in growing and non-growing cells: a physiological approach. *International Journal of Medical Microbiology, 295***,** 547-565.

Kolenbrander, P. (1993). Coaggregation of human oral bacteria: potential role in the accretion of dental plaque. *Journal of Applied Bacteriology, 74***,** 79S-86S.

Kolenbrander, P. E., Andersen, R. N., Baker, R. A. & Jenkinson, H. F. (1998). The Adhesion-Associated sca Operon in *Streptococcus gordonii* Encodes an Inducible High-Affinity ABC Transporter for Mn2+ Uptake. *Journal of Bacteriology, 180***,** 290-295.

Kondo, T. (2008). Tissue proteomics for cancer biomarker development-Laser microdissection and 2D-DIGE. *BMB Reports, 41***,** 626-634.

Kontinen, V. P. & Sarvas, M. (1993). The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion. *Molecular Microbiology, 8***,** 727-737.

Kosic, N., Sugai, M., Fan, C.-K. & Wu, H. (1993). Processing of lipid-modified prolipoprotein requires energy and sec gene products *in vivo*. *Journal of Bacteriology, 175***,** 6113-6117.

Köthe, M., Antl, M., Huber, B., Stoecker, K., Ebrecht, D., Steinmetz, I. & Eberl, L. (2003). Killing of *Caenorhabditis elegans* by *Burkholderia cepacia* is controlled by the cep quorum-sensing system. *Cellular Microbiology, 5***,** 343-351.

Kovacs-Simon, A., Titball, R. W. & Michell, S. L. (2011). Lipoproteins of bacterial pathogens. *Infection and Immunity, 79***,** 548-561.

Kuehnert, M. J., Kruszon-Moran, D., Hill, H. A., McQuillan, G., McAllister, S. K., Fosheim, G., McDougal, L. K., Chaitram, J., Jensen, B. & Fridkin, S. K. (2006).

Prevalence of *Staphylococcus aureus* nasal colonization in the United States, 2001–2002. *Journal of Infectious Diseases, 193***,** 172-179.

Kuklin, N. A., Clark, D. J., Secore, S., Cook, J., Cope, L. D., McNeely, T., Noble, L., Brown, M. J., Zorman, J. K. & Wang, X. M. (2006). A novel *Staphylococcus aureus* vaccine: iron surface determinant B induces rapid antibody responses in rhesus macaques and specific increased survival in a murine *S. aureus* sepsis model. *Infection and Immunity, 74***,** 2215-2223.

Kumar, H., Kawai, T. & Akira, S. (2011). Pathogen recognition by the innate immune system. *International Reviews of Immunology, 30***,** 16-34.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K.-i. & Nagai, Y. (2001). Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *The Lancet, 357***,** 1225-1240.

Kurokawa, K., Lee, H., Roh, K.-B., Asanuma, M., Kim, Y. S., Nakayama, H., Shiratsuchi, A., Choi, Y., Takeuchi, O. & Kang, H. J. (2009). The triacylated ATP binding cluster transporter substrate-binding lipoprotein of *Staphylococcus aureus* functions as a native ligand for Toll-like receptor 2. *Journal of Biological Chemistry, 284***,** 8406-8411.

Kurokawa, K., Ryu, K.-H., Ichikawa, R., Masuda, A., Kim, M.-S., Lee, H., Chae, J.-H., Shimizu, T., Saitoh, T. & Kuwano, K. (2012). Novel bacterial lipoprotein structures conserved in low-GC content Gram-positive bacteria are recognized by Toll-like receptor 2. *Journal of Biological Chemistry, 287***,** 13170-13181.

Kurokawa., Kim, M.-S., Ichikawa, R., Ryu, K.-H., Dohmae, N., Nakayama, H. & Lee, B. L. (2012). Environment-mediated accumulation of diacyl lipoproteins over their triacyl counterparts in *Staphylococcus aureus*. *Journal of Bacteriology, 194***,** 3299-3306.

Kurz, C. L. & Ewbank, J. J. (2000). *Caenorhabditis elegans* for the study of host–pathogen interactions. *Trends in Microbiology, 8***,** 142-144.

Labischinskia, H., Ehlerta, K. & Berger-Bächib, B. (1998). Leading articles The targeting of factors necessary for expression of methicillin resistance in staphylococci. *Journal of Antimicrobial Chemotherapy, 41***,** 581-584.

Ladhani, S. (2003). Understanding the mechanism of action of the exfoliative toxins of *Staphylococcus aureus*. *FEMS Immunology & Medical Microbiology, 39***,** 181-189.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9***,** 357-359.

Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & Lopez, R. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23***,** 2947-2948.

Larner-Svensson, H., Worning, P., Bartels, M. D., Hansen, L. H., Boye, K. & Westh, H. (2013). Complete genome sequence of *Staphylococcus aureus* strain M1, a unique t024-ST8-IVa Danish methicillin-resistant *S. aureus* clone. *Genome Announcements, 1***,** e00336-13.

Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I. R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penadés, J. R. & Valle, J. (2011).

Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences, 108***,** 20172-20177.

Laupland, K. B., Lyytikäinen, O., Søgaard, M., Kennedy, K., Knudsen, J., Ostergaard, C., Galbraith, J., Valiquette, L., Jacobsson, G. & Collignon, P. (2013). The changing epidemiology of *Staphylococcus aureus* bloodstream infection: a multinational population-based surveillance study. *Clinical Microbiology and Infection, 19***,** 465-471.

Lee, P. A., Tullman-Ercek, D. & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. *Annual Review of Microbiology, 60***,** 373.

Lee, S. G. & Fischetti, V. A. (2006). Purification and characterization of LPXTGase from *Staphylococcus aureus*: the amino acid composition mirrors that found in the peptidoglycan. *Journal of Bacteriology, 188***,** 389-398.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.-M. & Hoffmann, J. A. (1996). The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell, 86***,** 973-983.

Lettre, G. & Hengartner, M. O. (2006). Developmental apoptosis in *C. elegans*: a complex CEDnario. *Nature Reviews Molecular Cell Biology, 7***,** 97-108.

Levy, O. (1996). Antibiotic proteins of polymorphonuclear leukocytes. *European Journal of Haematology, 56***,** 263-277.

Lewis, J. A. & Fleming, J. (1995). Basic culture methods. *Caenorhabditis elegans: modern biological analysis of an organism, 48***,** 25-26.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27***,** 2987-2993.

Li, Y., Cao, B., Zhang, Y., Zhou, J., Yang, B. & Wang, L. (2011). Complete genome sequence of *Staphylococcus aureus* T0131, an ST239-MRSA-SCC*mec* type III clone isolated in China. *Journal of Bacteriology, 193***,** 3411-3412.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A. & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology, 11***,** 2453-2465.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25***,** 1451-1452.

Lim, H., Eng, J., Yates, J. R., Tollaksen, S. L., Giometti, C. S., Holden, J. F., Adams, M. W., Reich, C. I., Olsen, G. J. & Hays, L. G. (2003). Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide mass mapping to μ LC-ESI tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry, 14***,** 957-970.

Lindsay, J. A. & Holden, M. T. G. (2004). *Staphylococcus aureus* : superbug, super genome? *Trends in Microbiology, 12***,** 378-385.

Lindsay, J. A. & Holden, M. T. G. (2006). Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Functional & Integrative Genomics, 6***,** 186-201.

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D. & Hinds, J. (2006). Microarrays reveal that each of

the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *Journal of Bacteriology, 188***,** 669-676.

Lodish, H. F., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. & Darnell, J. 2000. *Molecular Cell Biology*, Citeseer.

Lowder, B. V., Guinane, C. M., Zakour, N. L. B., Weinert, L. A., Conway-Morris, A., Cartwright, R. A., Simpson, A. J., Rambaut, A., Nübel, U. & Fitzgerald, J. R. (2009). Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences, 106***,** 19545-19550.

Lowy, F. D. (1998). *Staphylococcus aureus* infections. *New England Journal of Medicine, 339***,** 520-532.

MacBeath, G. (2002). Protein microarrays and proteomics. *Nature Genetics, 32***,** 526-532.

Maduro, M. F. & Rothman, J. H. (2002). Making worm guts: the gene regulatory network of the *Caenorhabditis elegans* endoderm. *Developmental Biology, 246***,** 68-85.

Mallo, G. V., Kurz, C. L., Couillault, C., Pujol, N., Granjeaud, S., Kohara, Y. & Ewbank, J. J. (2002). Inducible antibacterial defense system in *C. elegans*. *Current Biology, 12***,** 1209-1214.

Mallorquí-Fernández, G., Marrero, A., García-Piquè, S., García-Castellanos, R. & Gomis-Rüth, F. X. (2004). Staphylococcal methicillin resistance: fine focus on folds and functions. *FEMS Microbiology Letters, 235***,** 1-8.

Mandatory Surveillance MRSA MRSA and MSSA bacteraemia and *C. difficile* infection mandatory data (official statistics). July 2014. *Public Health England, https://www.gov.uk/government/statistics/mrsa-mssa-and-e-coli-bacteraemia-and-c-difficile-infection-annual-epidemiological-commentary*.

Mann, M., Hendrickson, R. C. & Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry, 70***,** 437-473.

Mann, M. & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology, 21***,** 255-261.

Manna, A. C., Bayer, M. G. & Cheung, A. L. (1998). Transcriptional analysis of different promoters in the sar locus in *Staphylococcus aureus*. *Journal of Bacteriology, 180***,** 3828-3836.

Marie, B., Hernandez, D., Francois, P. & Schrenzel, J. (2010). New approaches for functional genomic studies in staphylococci. *International Journal of Medical Microbiology, 300***,** 88-97.

Mariotti, P., Malito, E., Biancucci, M., Lo Surdo, P., Mishra, R. P. N., Nardi-Dei, V., Savino, S., Nissum, M., Spraggon, G., Grandi, G., Bagnoli, F. & Bottomley, M. (2013). Structural and functional characterization of the *Staphylococcus aureus* virulence factor and vaccine candidate FhuD2. *Biochemical Journal, 449***,** 683-693.

Marrack, P. & Kappler, J. (1990). The staphylococcal enterotoxins and their relatives. *Science, 248***,** 705-711.

Marshall, C. & Spelman, D. (2007). Is throat screening necessary to detect methicillin-resistant *Staphylococcus aureus* colonization in patients upon admission to an intensive care unit? *Journal of Clinical Microbiology, 45***,** 3855-3855.

Mazmanian, S., Liu, G., Jensen, E. R., Lenoy, E. & Schneewind, O. (2000). *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proceedings of the National Academy of Sciences, 97***,** 5510-5515.

Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. (1999). *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science, 285***,** 760-763.

Mazmanian, S. K., Ton-That, H. & Schneewind, O. (2001). Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. *Molecular Microbiology, 40***,** 1049-1057.

McAleese, F., Wu, S. W., Sieradzki, K., Dunman, P., Murphy, E., Projan, S. & Tomasz, A. (2006). Overexpression of genes of the cell wall stimulon in clinical isolates of *Staphylococcus aureus* exhibiting vancomycin-intermediate-*S. aureus*-type resistance to vancomycin. *Journal of Bacteriology, 188***,** 1120-1133.

McCarthy, A. J. & Lindsay, J. A. (2010). Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC Microbiology, 10***,** 173.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., Vanderpool, C. K. & Tjaden, B. (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research, 41***,** e140-e140.

McCullers, J. A. (2006). Insights into the interaction between influenza virus and pneumococcus. *Clinical Microbiology Reviews, 19***,** 571-582.

Mcdougal, L. K. & Thornsberry, C. (1986). The role of beta-lactamase in staphylococcal resistance to penicillinase-resistant penicillins and cephalosporins. *Journal of Clinical Microbiology, 23***,** 832-839.

Meyer, E., Aglyamova, G. V., Wang, S., Buchanan-Carter, J., Abrego, D., Colbourne, J. K., Willis, B. L. & Matz, M. V. (2009). Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics, 10***,** 219.

Millet, A. C. & Ewbank, J. J. (2004). Immunity in *Caenorhabditis elegans*. *Current Opinion In Immunology, 16***,** 4-9.

Mishra, R. P. N., Mariotti, P., Fiaschi, L., Nosari, S., MacCari, S., Liberatori, S., Fontana, M. R., Pezzicoli, A., De Falco, M. G., Falugi, F., Altindis, E., Serruto, D., Grandi, G. & Bagnoli, F. (2012). *Staphylococcus aureus* FhuD2 is involved in the early phase of staphylococcal dissemination and generates protective immunity in mice. *Journal of Infectious Diseases, 206***,** 1041-1049.

Moore, P. & Lindsay, J. (2002). Molecular characterisation of the dominant UK methicillin-resistant *Staphylococcus aureus* strains, EMRSA-15 and EMRSA-16. *Journal of Medical Microbiology, 51***,** 516-521.

Moosdeen, F., Williams, J. & Secker, A. (1988). Standardization of inoculum size for disc susceptibility testing: a preliminary report of a spectrophotometric method. *Journal of Antimicrobial Chemotherapy, 21***,** 439-443.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods, 5***,** 621-628.

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American, 262***,** 56-61.

Muñoz-Planillo, R., Franchi, L., Miller, L. S. & Núñez, G. (2009). A critical role for hemolysins and bacterial lipoproteins in *Staphylococcus aureus* induced activation of the Nlrp3 inflammasome. *The Journal of Immunology, 183***,** 3942-3948.

Mylonakis, E., Ausubel, F. M., Perfect, J. R., Heitman, J. & Calderwood, S. B. (2002). Killing of *Caenorhabditis elegan*s by *Cryptococcus neoformans* as a model of yeast pathogenesis. *Proceedings of the National Academy of Sciences, 99***,** 15675-15680.

Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., Gill, S., Francois, P. & Cheung, A. L. (2011). Whole-genome sequencing of *Staphylococcus aureus* strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. *Journal of Bacteriology, 193***,** 2332-2335.

Nakayama, H., Kurokawa, K. & Lee, B. L. (2012). Lipoproteins in bacteria: structures and biosynthetic pathways. *FEBS Journal, 279***,** 4247-4268.

Nandakumar, R., Nandakumar, M., Marten, M. R. & Ross, J. M. (2005). Proteome analysis of membrane and cell wall associated proteins from *Staphylococcus aureus*. *Journal of Proteome Research, 4***,** 250-257.

Natale, P., Brüser, T. & Driessen, A. J. (2008). Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1778***,** 1735-1756.

Navarre, W. W. & Schneewind, O. (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews, 63***,** 174-229.

Neidig, N., Paul, R. J., Scheu, S. & Jousset, A. (2011). Secondary metabolites of *Pseudomonas fluorescens* CHA0 drive complex non-trophic interactions with bacterivorous nematodes. *Microbial Ecology, 61***,** 853-859.

Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., Van Sluyter, S. C. & Haynes, P. A. (2011). Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics, 11***,** 535-553.

Neoh, H.-m., Cui, L., Yuzawa, H., Takeuchi, F., Matsuo, M. & Hiramatsu, K. (2008). Mutated response regulator *graR* is responsible for phenotypic conversion of *Staphylococcus aureus* from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance. *Antimicrobial Agents and Chemotherapy, 52***,** 45-53.

Neufert, C., Pai, R. K., Noss, E. H., Berger, M., Boom, W. H. & Harding, C. V. (2001). *Mycobacterium tuberculosis* 19-kDa lipoprotein promotes neutrophil activation. *The Journal of Immunology, 167***,** 1542-1549.

Nguyen, M. T., Kraft, B., Yu, W., Demicrioglu, D. D., Hertlein, T., Burian, M., Schmaler, M., Boller, K., Bekeredjian-Ding, I. & Ohlsen, K. (2015). The νSaα Specific Lipoprotein Like Cluster (*lpl*) of *S. aureus* USA300 Contributes to Immune Stimulation and Invasion in Human Cells. *PLoS Pathogens, 11*.

Nicholas, H. R. & Hodgkin, J. (2004a). The ERK MAP kinase cascade mediates tail swelling and a protective response to rectal infection in *C. elegans*. *Current Biology, 14***,** 1256-1261.

Nicholas, H. R. & Hodgkin, J. (2004b). Responses to infection and possible recognition strategies in the innate immune system of *Caenorhabditis elegans*. *Molecular Immunology, 41***,** 479-493.

Nielsen, J. S., Christiansen, M. H., Bonde, M., Gottschalk, S., Frees, D., Thomsen, L. E. & Kallipolitis, B. H. (2011). Searching for small σB-regulated genes in *Staphylococcus aureus*. *Archives of Microbiology, 193***,** 23-34.

Nilsson, P. & Ripa, T. (2006). *Staphylococcus aureus* throat colonization is more frequent than colonization in the anterior nares. *Journal Of Clinical Microbiology, 44***,** 3334-3339.

Noguchi, N., Nakaminami, H., Nishijima, S., Kurokawa, I., So, H. & Sasatsu, M. (2006). Antimicrobial agent of susceptibilities and antiseptic resistance gene distribution among methicillin-resistant *Staphylococcus aureus* isolates from patients with impetigo and staphylococcal scalded skin syndrome. *Journal of Clinical Microbiology, 44***,** 2119-2125.

Novick, R. P. (2003). Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Molecular Microbiology, 48***,** 1429-1449.

Novick, R. P., Ross, H., Projan, S., Kornblum, J., Kreiswirth, B. & Moghazeh, S. (1993). Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *The EMBO journal, 12***,** 3967.

O'Neill, E., Pozzi, C., Houston, P., Humphreys, H., Robinson, D. A., Loughman, A., Foster, T. J. & O'Gara, J. P. (2008). A novel *Staphylococcus aureus* biofilm phenotype mediated by the fibronectin-binding proteins, FnBPA and FnBPB. *Journal of Bacteriology, 190***,** 3835-3850.

O'quinn, A., Wiegand, E. & Jeddeloh, J. (2001). *Burkholderia pseudomallei* kills the nematode *Caenorhabditis elegans* using an endotoxin-mediated paralysis. *Cellular Microbiology, 3***,** 381-393.

O'Rourke, D., Baban, D., Demidova, M., Mott, R. & Hodgkin, J. (2006). Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. *Genome Research, 16***,** 1005-1016.

Ohta, T., Hirakawa, H., Morikawa, K., Maruyama, A., Inose, Y., Yamashita, A., Oshima, K., Kuroda, M., Hattori, M. & Hiramatsu, K. (2004). Nucleotide substitutions in *Staphylococcus aureus* strains, Mu50, Mu3, and N315. *DNA Research, 11***,** 51-56.

Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A. & Ahn, N. G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & Cellular Proteomics, 4***,** 1487-1502.

Osanai, A., Hu, D.-L. & Nakane, A. (2012). *Caenorhabditis elegans* avoids staphylococcal superantigenic toxins via 5-hydroxytryptamine-dependent pathway. *Canadian Journal Of Microbiology, 58***,** 1268-1277.

Oshlack, A., Robinson, M. D. & Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biol, 11***,** 220.

Osterholm, M. T. & Forfang, J. C. (1982). Toxic-shock syndrome in Minnesota: results of an active-passive surveillance system. *Journal of Infectious Diseases, 145***,** 458-464.

Otto, M. (2010). Basis of virulence in community-associated methicillin-resistant *Staphylococcus aureus*. *Annual Review of Microbiology, 64***,** 143-162.

Page, A. & Johnstone, I. (2007). The Cuticle in Worm Book (The *C. elegans* Research Community).

Pagels, M., Fuchs, S., Pané-Farré, J., Kohler, C., Menschner, L., Hecker, M., McNamarra, P. J., Bauer, M. C., Von Wachenfeldt, C. & Liebeke, M. (2010). Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in *Staphylococcus aureus*. *Molecular microbiology, 76***,** 1142-1161.

Paget, M. & Helmann, J. D. (2003). The sigma70 family of sigma factors. *Genome Biology, 4***,** 203.

Patti, J. M. (2011). Will we ever see the approval of a *Staphylococcus aureus* vaccine? *Expert Review of Anti-Infective Therapy, 9***,** 845-846.

Peng, H., Novick, R., Kreiswirth, B., Kornblum, J. & Schlievert, P. (1988). Cloning, characterization, and sequencing of an accessory gene regulator (agr) in *Staphylococcus aureus*. *Journal of Bacteriology, 170***,** 4365-4372.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research, 29***,** 45-49.

Pfaffl, M. W. (2004). Quantification strategies in real-time PCR. *AZ of Quantitative PCR, 1***,** 89-113.

Pierce, S. B., Costa, M., Wisotzkey, R., Devadhar, S., Homburger, S. A., Buchman, A. R., Ferguson, K. C., Heller, J., Platt, D. M. & Pasquinelli, A. A. (2001). Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes & Development, 15***,** 672-686.

Pisetsky, D. S. (2007). The role of nuclear macromolecules in innate immunity. *Proceedings of the American Thoracic Society, 4***,** 258-262.

Pottumarthy, S., Fritsche, T. R. & Jones, R. N. (2005). Evaluation of alternative disk diffusion methods for detecting *mecA*-mediated oxacillin resistance in an international collection of staphylococci: validation report from the SENTRY Antimicrobial Surveillance Program. *Diagnostic Microbiology and Infectious Disease, 51***,** 57-62.

Powell, J. R. & Ausubel, F. M. 2008. Models of *Caenorhabditis elegans* infection by bacterial and fungal pathogens. *Innate Immunity.* Springer.

Pradel, E., Zhang, Y., Pujol, N., Matsuyama, T., Bargmann, C. I. & Ewbank, J. J. (2007). Detection and avoidance of a natural product from the pathogenic bacterium *Serratia marcescens* by *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences, 104***,** 2295-2300.

Pragman, A. A. & Schlievert, P. M. (2004). Virulence regulation in *Staphylococcus aureus*: the need for *in vivo* analysis of virulence factor regulation. *FEMS Immunology & Medical Microbiology, 42***,** 147-154.

Projan, S. & Novick, R. (1997). The molecular basis of pathogenicity. *The staphylococci in human disease. Churchill Livingstone, New York, NY***,** 55-81.

Pujol, N., Link, E. M., Liu, L. X., Kurz, C. L., Alloing, G., Tan, M.-W., Ray, K. P., Solari, R., Johnson, C. D. & Ewbank, J. J. (2001). A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. *Current Biology, 11***,** 809-821.

Qu, T., Feng, Y., Jiang, Y., Zhu, P., Wei, Z., Chen, Y., Otto, M. & Yu, Y. (2014). Whole genome analysis of a community-associated methicillin-resistant Staphylococcus aureus ST59 isolate from a case of human sepsis and severe pneumonia in China. *PLoS ONE, 9*.

Rabilloud, T., Vaezzadeh, A. R., Potier, N., Lelong, C., Leize-Wagner, E. & Chevallet, M. (2009). Power and limitations of electrophoretic separations in proteomics strategies. *Mass Spectrometry Reviews, 28***,** 816-843.

Rahman, O., Cummings, S. P., Harrington, D. J. & Sutcliffe, I. C. (2008). Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World Journal of Microbiology and Biotechnology, 24***,** 2377-2382.

Ravipaty, S. & Reilly, J. P. (2010). Comprehensive characterization of methicillin-resistant *Staphylococcus aureus* subsp. *S. aureus* COL secretome by two-dimensional liquid chromatography and mass spectrometry. *Molecular & Cellular Proteomics, 9***,** 1898-1919.

Raymond, E. & Traub, W. (1970). Identification of staphylococci isolated from clinical material. *Applied Microbiology, 19***,** 919-922.

Réglier-Poupet, H., Frehel, C., Dubail, I., Beretti, J.-L., Berche, P., Charbit, A. & Raynaud, C. (2003). Maturation of lipoproteins by type II signal peptidase is required for phagosomal escape of *Listeria monocytogenes*. *Journal of Biological Chemistry, 278***,** 49469-49477.

Reynolds, P. E. & Brown, D. F. (1985). Penicillin-binding proteins of β-lactam-resistant strains of *Staphylococcus aureus*: effect of growth conditions. *FEBS Letters, 192***,** 28-32.

Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R. (1997). The Primary Sex-Determination Signal and its Gene Target.

Roberts, C., Anderson, K. L., Murphy, E., Projan, S. J., Mounts, W., Hurlburt, B., Smeltzer, M., Overbeek, R., Disz, T. & Dunman, P. M. (2006). Characterizing the effect of the *Staphylococcus aureus* virulence factor regulator, *SarA*, on log-phase mRNA half-lives. *Journal of Bacteriology, 188***,** 2593-2603.

Robertson, H. M. & Thomas, J. H. (2006). The putative chemoreceptor families of *C. elegans*.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26***,** 139-140.

Roche, F. M., Massey, R., Peacock, S. J., Day, N. P. J., Visai, L., Speziale, P., Lam, A., Pallen, M. & Foster, T. J. (2003). Characterization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. *Microbiology, 149***,** 643-654.

Roeder, T., Stanisak, M., Gelhaus, C., Bruchhaus, I., Grötzinger, J. & Leippe, M. (2010). Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. *Developmental & Comparative Immunology, 34***,** 203-209.

Said-Salim, B., Dunman, P., McAleese, F., Macapagal, D., Murphy, E., McNamara, P., Arvidson, S., Foster, T., Projan, S. & Kreiswirth, B. (2003). Global regulation of *Staphylococcus aureus* genes by Rot. *Journal of Bacteriology, 185***,** 610-619.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4***,** 406-425.

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A. & Moreillon, P. (2011). *Staphylococcus aureus* host range and human-bovine host shift. *Applied and Environmental Microbiology, 77***,** 5908-5915.

Sakwinska, O., Kuhn, G., Balmelli, C., Francioli, P., Giddey, M., Perreten, V., Riesen, A., Zysset, F., Blanc, D. S. & Moreillon, P. (2009). Genetic diversity and ecological success of *Staphylococcus aureus* strains colonizing humans. *Applied and Environmental Microbiology, 75***,** 175-183.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74***,** 5463-5467.

Sankaran, K. & Wu, H. C. 1993. Bacterial lipoproteins. *Lipid modifications of proteins.* CRC Press Florida.

Sankaran, K. & Wu, H. C. (1994). Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *Journal of Biological Chemistry, 269***,** 19701-19706.

Sankaran, K. & Wu, H. C. (1995). Bacterial prolipoprotein signal peptidase. *Methods in Enzymology, 248***,** 169-180.

Sass, P., Berscheid, A., Jansen, A., Oedenkoven, M., Szekat, C., Strittmatter, A., Gottschalk, G. & Bierbaum, G. (2012). Genome sequence of *Staphylococcus aureus* VC40, a vancomycin-and daptomycin-resistant strain, to study the genetics of development of resistance to currently applied last-resort antibiotics. *Journal of Bacteriology, 194***,** 2107-2108.

Schaumburg, F., Alabi, A., Peters, G. & Becker, K. (2014). New epidemiology of *Staphylococcus aureus* infection in Africa. *Clinical Microbiology and Infection, 20***,** 589-596.

Schijffelen, M., Konstantinov, S. R., Lina, G., Spiliopoulou, I., van Duijkeren, E., Brouwer, E. C. & Fluit, A. C. (2013). Whole Genome Analysis of Epidemiologically Closely Related *Staphylococcus aureus* Isolates. *PLoS ONE, 8*.

Schlag, S., Fuchs, S., Nerz, C., Gaupp, R., Engelmann, S., Liebeke, M., Lalk, M., Hecker, M. & Götz, F. (2008). Characterization of the oxygen-responsive NreABC regulon of *Staphylococcus aureus*. *Journal of Bacteriology, 190***,** 7847-7858.

Schmaler, M., Jann, N. J., Ferracin, F., Landolt, L. Z., Biswas, L., Götz, F. & Landmann, R. (2009). Lipoproteins in *Staphylococcus aureus* mediate inflammation by TLR2 and iron-dependent growth *in vivo*. *The Journal of Immunology, 182***,** 7110-7118.

Schmaler, M., Jann, N. J., Götz, F. & Landmann, R. (2010). Staphylococcal lipoproteins and their role in bacterial survival in mice. *International Journal of Medical Microbiology, 300***,** 155-160.

Schmidt, F., Scharf, S. S., Hildebrandt, P., Burian, M., Bernhardt, J., Dhople, V., Kalinka, J., Gutjahr, M., Hammer, E. & Völker, U. (2010). Time-resolved quantitative proteome profiling of host–pathogen interactions: The response of *Staphylococcus aureus* RN1HG to internalisation by human airway epithelial cells. *Proteomics, 10***,** 2801-2811.

Schmittgen, T. D. & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols, 3***,** 1101-1108.

Schneewind, O., Mihaylova-Petkov, D. & Model, P. (1993). Cell wall sorting signals in surface proteins of gram-positive bacteria. *The EMBO Journal, 12***,** 4803.

Schnellmann, C., Gerber, V., Rossano, A., Jaquier, V., Panchaud, Y., Doherr, M. G., Thomann, A., Straub, R. & Perreten, V. (2006). Presence of new mecA and mph (C) variants conferring antibiotic resistance in Staphylococcus spp. isolated from the skin of horses before and after clinic admission. *Journal of Clinical Microbiology, 44***,** 4444-4454.

Schulenburg, H. & Boehnisch, C. (2008). Diversification and adaptive sequence evolution of *Caenorhabditis* lysozymes (Nematoda: Rhabditidae). *BMC Evolutionary Biology, 8***,** 114.

Schulenburg, H., Hoeppner, M. P., Weiner, J. & Bornberg-Bauer, E. (2008). Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. *Immunobiology, 213***,** 237-250.

Schulenburg, H., Léopold Kurz, C. & Ewbank, J. J. (2004). Evolution of the innate immune system: the worm perspective. *Immunological Reviews, 198***,** 36-58.

Schumacher, M. A., Hurlburt, B. K. & Brennan, R. G. (2001). Crystal structures of *SarA*, a pleiotropic regulator of virulence genes in *S. aureus*. *Nature, 409***,** 215-219.

Schwan, W. R., Coulter, S. N., Ng, E. Y., Langhorne, M. H., Ritchie, H. D., Brody, L. L., Westbrock-Wadman, S., Bayer, A. S., Folger, K. R. & Stover, C. K. (1998). Identification and characterization of the PutP proline permease that contributes to *in vivo* survival of *Staphylococcus aureus* in animal models. *Infection and Immunity, 66***,** 567-572.

Scott, E., Duty, S. & Callahan, M. (2008). A pilot study to isolate *Staphylococcus aureus* and methicillin-resistant *S aureus* from environmental surfaces in the home. *American Journal of Infection Control, 36***,** 458-460.

Sebulsky, M. T. & Heinrichs, D. E. (2001). Identification and characterization of *fhuD1* and *fhuD2*, two genes involved in iron-hydroxamate uptake in *Staphylococcus aureus*. *Journal of Bacteriology, 183***,** 4994-5000.

Sebulsky, M. T., Hohnstein, D., Hunter, M. D. & Heinrichs, D. E. (2000). Identification and Characterization of a Membrane Permease Involved in Iron Hydroxamate Transport in *Staphylococcus aureus*. *Journal of Bacteriology, 182***,** 4394-4400.

Sebulsky, M. T., Speziali, C. D., Shilton, B. H., Edgell, D. R. & Heinrichs, D. E. (2004). FhuD1, a Ferric Hydroxamate-binding Lipoprotein in *Staphylococcus aureus* a case of gene duplicatin and lateral transfer *Journal of Biological Chemistry, 279***,** 53152-53159.

Serebryakova, M. V., Demina, I. A., Galyamina, M. A., Kondratov, I. G., Ladygina, V. G. & Govorun, V. M. (2011). The acylation state of surface lipoproteins of mollicute *Acholeplasma laidlawii*. *Journal of Biological Chemistry, 286***,** 22769-22776.

Setubal, J. C., Reis, M., Matsunaga, J. & Haake, D. A. (2006). Lipoprotein computational prediction in spirochaetal genomes. *Microbiology, 152***,** 113-121.

Shapira, M., Hamlin, B. J., Rong, J., Chen, K., Ronen, M. & Tan, M.-W. (2006). A conserved role for a GATA transcription factor in regulating epithelial innate immune responses. *Proceedings of the National Academy of Sciences, 103***,** 14086-14091.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J. & Reinhardt, R. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature, 464***,** 250-255.

Sheldon, J. R. & Heinrichs, D. E. (2012). The iron-regulated staphylococcal lipoproteins. *Frontiers in Cellular and Infection Microbiology, 2*.

Shibata, K.-i., Hasebe, A., Into, T., Yamada, M. & Watanabe, T. (2000). The N-terminal lipopeptide of a 44-kDa membrane-bound lipoprotein of *Mycoplasma salivarium* is responsible for the expression of intercellular adhesion molecule-1 on the cell surface of normal human gingival fibroblasts. *The Journal of Immunology, 165***,** 6538-6544.

Shivers, R. P., Kooistra, T., Chu, S. W., Pagano, D. J. & Kim, D. H. (2009). Tissue-specific activities of an immune signaling module regulate physiological responses to pathogenic and nutritional bacteria in *C. elegans*. *Cell Host & Microbe, 6***,** 321-330.

Shockman, G. D. & Barren, J. (1983). Structure, function, and assembly of cell walls of gram-positive bacteria. *Annual Reviews in Microbiology, 37***,** 501-527.

Sibbald, M., Ziebandt, A., Engelmann, S., Hecker, M., De Jong, A., Harmsen, H., Raangs, G., Stokroos, I., Arends, J. & Dubois, J. (2006). Mapping the pathways to staphylococcal pathogenesis by comparative secretomics. *Microbiology and Molecular Biology Reviews, 70***,** 755-788.

Sifri, C. D., Begun, J., Ausubel, F. M. & Calderwood, S. B. (2003). *Caenorhabditis elegans* as a model host for *Staphylococcus aureus* pathogenesis. *Infection and Immunity, 71***,** 2208-2217.

Sinha, B., François, P. P., Nüße, O., Foti, M., Hartford, O. M., Vaudaux, P., Foster, T. J., Lew, D. P., Herrmann, M. & Krause, K. H. (1999). Fibronectin-binding protein acts as *Staphylococcus aureus* invasin via fibronectin bridging to integrin α5β1. *Cellular Microbiology, 1***,** 101-117.

Skov, R., Christiansen, K., Dancer, S. J., Daum, R. S., Dryden, M., Huang, Y.-C. & Lowy, F. D. (2012). Update on the prevention and control of community-acquired meticillin-resistant *Staphylococcus aureus* (CA-MRSA). *International Journal of Antimicrobial Agents, 39***,** 193-200.

Solovyev, V. & Salamov, A. (2011). Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics And Its Applications In Agriculture, Biomedicine And Environmental Studies***,** 61-78.

Stoll, H., Dengjel, J., Nerz, C. & Götz, F. (2005). *Staphylococcus aureus* deficient in lipidation of prelipoproteins is attenuated in growth and immune activation. *Infection and Immunity, 73***,** 2411-2423.

Strange, K. (2003). From genes to integrative physiology: ion channel and transporter biology in *Caenorhabditis elegans*. *Physiological Reviews, 83***,** 377-415.

Stranger-Jones, Y. K., Bae, T. & Schneewind, O. (2006). Vaccine assembly from surface proteins of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences, 103***,** 16942-16947.

Styers, D., Sheehan, D. J., Hogan, P. & Sahm, D. F. (2006). Laboratory-based surveillance of current antimicrobial resistance patterns and trends among *Staphylococcus aureus*: 2005 status in the United States. *Annals of Clinical Microbiology and Antimicrobials, 5***,** 2.

Sugai, M. & Wu, H. (1992). Export of the outer membrane lipoprotein is defective in secD, secE, and secF mutants of *Escherichia coli*. *Journal of Bacteriology, 174***,** 2511-2516.

Sulston, J. E. & Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology, 56***,** 110-156.

Sung, J. M. L., Lloyd, D. H. & Lindsay, J. A. (2008). *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology, 154***,** 1949-1959.

Sutcliffe, I. C. & Harrington, D. J. (2002). Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. *Microbiology, 148***,** 2065-2077.

Sutcliffe, I. C. & Russell, R. (1995). Lipoproteins of gram-positive bacteria. *Journal of Bacteriology, 177***,** 1123.

Sydnor, E. R. & Perl, T. M. (2011). Hospital epidemiology and infection control in acute-care settings. *Clinical Microbiology Reviews, 24***,** 141-173.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123***,** 585-595.

Takeuchi, O., Hoshino, K. & Akira, S. (2000). Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to *Staphylococcus aureus* infection. *The Journal of Immunology, 165***,** 5392-5396.

Tam, R. & Saier, M. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological Reviews, 57***,** 320-346.

Tan, M.-W., Mahajan-Miklos, S. & Ausubel, F. M. (1999). Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proceedings of the National Academy of Sciences, 96***,** 715-720.

Tanaka, K., Matsuyama, S.-I. & Tokuda, H. (2001). Deletion of lolB, encoding an outer membrane lipoprotein, is lethal for *Escherichia coli* and causes accumulation of

lipoprotein localization intermediates in the periplasm. *Journal of Bacteriology, 183***,** 6538-6542.

Tegmark, K., Karlsson, A. & Arvidson, S. (2000). Identification and characterization of SarH1, a new global regulator of virulence gene expression in *Staphylococcus aureus*. *Molecular Microbiology, 37***,** 398-409.

Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R. & Grandi, G. (2006). Pili in gram-positive pathogens. *Nature Reviews Microbiology, 4***,** 509-519.

Terada, M., Kuroda, T., Matsuyama, S.-i. & Tokuda, H. (2001). Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of *Escherichia coli*. *Journal of Biological Chemistry, 276***,** 47690-47694.

Thakker, M., Park, J. S., Carey, V. & Lee, J. C. (1998). *Staphylococcus aureus* serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. *Infection and Immunity, 66***,** 5183-5189.

Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J. & Mann, M. (2011). Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular & Cellular Proteomics, 10***,** M110. 003699.

Thomas, J. H. (2006). Concerted evolution of two novel protein families in Caenorhabditis species. *Genetics, 172***,** 2269-2281.

Thompson, B. J., Widdick, D. A., Hicks, M. G., Chandra, G., Sutcliffe, I. C., Palmer, T. & Hutchings, M. I. (2010). Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium *Streptomyces coelicolor*. *Molecular Microbiology, 77***,** 943-957.

Ting, J. P.-Y., Willingham, S. B. & Bergstralh, D. T. (2008). NLRs at the intersection of cell death and immunity. *Nature Reviews Immunology, 8***,** 372-379.

Tokajian, S. (2014). New epidemiology of *Staphylococcus aureus* infections in the Middle East. *Clinical Microbiology and Infection, 20***,** 624-628.

Tokunaga, M., Tokunaga, H. & Wu, H. C. (1982). Post-translational modification and processing of *Escherichia coli* prolipoprotein *in vitro*. *Proceedings of the National Academy of Sciences, 79***,** 2255-2259.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols, 7***,** 562-578.

Trivier, D. & Courcol, R. (1996). Iron depletion and virulence in *Staphylococcus aureus*. *FEMS Microbiology Letters, 141***,** 117-127.

Tschumi, A., Nai, C., Auchli, Y., Hunziker, P., Gehrig, P., Keller, P., Grau, T. & Sander, P. (2009). Identification of apolipoprotein N-acyltransferase (Lnt) in mycobacteria. *Journal of Biological Chemistry, 284***,** 27146-27156.

Tseng, C. W., Zhang, S. & Stewart, G. C. (2004). Accessory gene regulator control of staphyloccoccal enterotoxin d gene expression. *Journal of Bacteriology, 186***,** 1793-1801.

Twumasi-Boateng, K. & Shapira, M. (2012). Dissociation of immune responses from pathogen colonization supports pattern recognition in *C. elegans*. *PloS One, 7***,** e35400.

Ubukata, K., Yamashita, N. & Konno, M. (1985). Occurrence of a beta-lactam-inducible penicillin-binding protein in methicillin-resistant staphylococci. *Antimicrobial Agents and Chemotherapy, 27***,** 851-857.

Uhlemann, A.-C., Knox, J., Miller, M., Hafer, C., Vasquez, G., Ryan, M., Vavagiakis, P., Shi, Q. & Lowy, F. D. (2011). The environment as an unrecognized reservoir for community-associated methicillin resistant *Staphylococcus aureus* USA300: a case-control study. *PLoS One, 6***,** e22407.

Utaida, S., Dunman, P., Macapagal, D., Murphy, E., Projan, S., Singh, V., Jayaswal, R. & Wilkinson, B. (2003). Genome-wide transcriptional profiling of the response of *Staphylococcus aureus* to cell-wall-active antibiotics reveals a cell-wall-stress stimulon. *Microbiology, 149***,** 2719-2732.

Vidal-Ingigliardi, D., Lewenza, S. & Buddelmeijer, N. (2007). Identification of essential residues in apolipoprotein N-acyl transferase, a member of the CN hydrolase family. *Journal of Bacteriology, 189***,** 4456-4464.

Vlach, K. D., Boles, J. W. & Stiles, B. G. (2000). Telemetric evaluation of body temperature and physical activity as predictors of mortality in a murine model of staphylococcal enterotoxic shock. *Comparative Medicine, 50***,** 160-166.

Vogel, J. & Sharma, C. M. (2005). How to find small non-coding RNAs in bacteria. *Biological Chemistry, 386***,** 1219-1238.

Waldvogel, F. A. (1995). *Staphylococcus aureus* (including toxic shock syndrome). *Principles and practice of infectious diseases. Churchill Livingstone, New York, NY***,** 1754-1777.

Wang, Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10***,** 57-63.

Wang, J. & Kim, S. K. (2003). Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development, 130***,** 1621-1634.

Wang, J., Kodali, S., Lee, S. H., Galgoci, A., Painter, R., Dorso, K., Racine, F., Motyl, M., Hernandez, L. & Tinney, E. (2007). Discovery of platencin, a dual FabF and FabH inhibitor with *in vivo* antibiotic properties. *Proceedings of the National Academy of Sciences, 104***,** 7612-7616.

Ward, S. & Lamunyon, C. (1995). Sperm precedence in a hermaphroditic nematode (*Caenorhabditis elegans*) is due to competitive superiority of male sperm. *Experientia, 58***,** 817-823.

Wardenburg, J. B., Williams, W. A. & Missiakas, D. (2006). Host defenses against *Staphylococcus aureus* infection require recognition of bacterial lipoproteins. *Proceedings of the National Academy of Sciences, 103***,** 13831-13836.

Werck-Reichhart, D. & Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biol, 1***,** 3003.1-3003.9.

Wertheim, H. F., Melles, D. C., Vos, M. C., van Leeuwen, W., van Belkum, A., Verbrugh, H. A. & Nouwen, J. L. (2005). The role of nasal carriage in *Staphylococcus aureus* infections. *The Lancet Infectious Diseases, 5***,** 751-762.

Wesley, K. & Bannerman, T. L. (1994). Update on clinical significance of coagulase-negative staphylococci. *Clinical Microbiology Reviews, 7***,** 117-140.

Wilkinson, B. 1997. The Staphylococci in Human Diseases. London: Churchill Livingston.

Williams, W. A., Zhang, R.-g., Zhou, M., Joachimiak, G., Gornicki, P., Missiakas, D. & Joachimiak, A. (2004). The membrane-associated lipoprotein-9 GmpC from *Staphylococcus aureus* binds the dipeptide GlyMet via side chain interactions. *Biochemistry, 43***,** 16193-16202.

Williamson, D., Coombs, G. & Nimmo, G. (2014). *Staphylococcus aureus* 'Down Under': contemporary epidemiology of *S. aureus* in Australia, New Zealand, and the South West Pacific. *Clinical Microbiology and Infection, 20***,** 597-604.

Willis, A., O'CONNOR, J. J. & Smith, J. (1966). Some observations on staphylococcal pigmentation.

Windle, H. J., Brown, P. A. & Kelleher, D. P. (2010). Proteomics of bacterial pathogenicity: therapeutic implications. *Proteomics-Clinical Applications, 4***,** 215-227.

Witney, A. A., Marsden, G. L., Holden, M. T. G., Stabler, R. A., Husain, S. E., Vass, J. K., Butcher, P. D., Hinds, J. & Lindsay, J. A. (2005). Design, validation, and application of a seven-strain *Staphylococcus aureus* PCR product microarray for comparative genomics. *Applied and Environmental Microbiology, 71***,** 7504-7514.

Witte, W., Pasemann, B. & Cuny, C. (2007a). Detection of low-level oxacillin resistance in *mecA*-positive *Staphylococcus aureus*. *Clinical Microbiology and Infection, 13***,** 408-412.

Witte, W., Strommenger, B., Stanek, C. & Cuny, C. (2007b). Methicillin-resistant *Staphylococcus aureus* ST398 in humans and animals, Central Europe. *Emerging Infectious Diseases, 13***,** 255.

Wong, D., Bazopoulou, D., Pujol, N., Tavernarakis, N. & Ewbank, J. J. (2007). Genome-wide investigation reveals pathogen-specific and shared signatures in the response of *Caenorhabditis elegans* to infection. *Genome Biol, 8***,** R194.

Wood, W. B. 1988. 1 Introduction to *C. elegans* Biology. *Cold Spring Harbor Monograph Archive.*

Wright, J. A. & Nair, S. P. (2012). The lipoprotein components of the Isd and Hts transport systems are dispensable for acquisition of heme by *Staphylococcus aureus*. *FEMS Microbiology Letters, 329***,** 177-185.

Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B. A. & Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. *Genome Research, 20***,** 133-141.

Yates, J. R., Ruse, C. I. & Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Review of Biomedical Engineering, 11***,** 49-79.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13***,** 134.

Yeung, Y.-G., Nieves, E., Angeletti, R. H. & Stanley, E. R. (2008). Removal of detergents from protein digests for mass spectrometry analysis. *Analytical Biochemistry, 382***,** 135-137.

Zhang, H., Niesel, D. W., Peterson, J. W. & Klimpel, G. R. (1998). Lipoprotein release by bacteria: potential factor in bacterial pathogenesis. *Infection and Immunity, 66***,** 5196-5201.

Zhang, L., Yu, J., Xie, Y., Lin, H., Huang, Z., Xu, L., Gelbič, I. & Guan, X. (2014). Biological Activity of *Bacillus thuringiensis* (Bacillales: Bacillaceae) Chitinase Against *Caenorhabditis elegans* (Rhabditida: Rhabditidae). *Journal of Economic Entomology, 107***,** 551-558.

Ziebandt, A. K., Becher, D., Ohlsen, K., Hacker, J., Hecker, M. & Engelmann, S. (2004). The influence of *agr* and σB in growth phase dependent regulation of virulence factors in *Staphylococcus aureus*. *Proteomics, 4***,** 3034-3047.

Zückert, W. R. (2014). Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1843***,** 1509-1516.

Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L. & Washburn, M. P. (2006). Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces c erevisiae*. *Journal of Proteome Research, 5***,** 2339-2347.

Appendices

Appendix 1

Table 1. *SAR0443* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0443	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1480	100%	0.0	100%	801/801
S. aureus RN4282	1153	90%	0.0	99%	699/699
S. aureus T1	1480	100%	0.0	100%	801/801
S. aureus COL	1133	99%	0.0	96%	669/697
S. aureus MSSA476	1153	99%	0.0	97%	668/690
S. aureus MW2	1153	99%	0.0	97%	668/690
S. aureus N315	1125	100%	0.0	94%	755/801
S. aureus Mu50	1225	100%	0.0	94%	755/801
S. aureus T0131	1133	99%	0.0	96%	669/697
S. aureus TW20	344	78%	0.0	77%	492/638
S. aureus BMB9393	933	99%	0.0	88%	698/794
S. aureus LGA251	1149	87%	0.0	96%	672/697
S. aureus RF122	678	90%	0.0	86%	539/625
S. aureus ST398	1153	99%	0.0	97%	668/690
S. aureus M013	601	90%	0.0	87%	469/541
S. aureus M1	1133	99%	0.0	96%	669/697
S. aureus Mu3	1225	100%	0.0	94%	755/801
S. aureus str. Newman	1133	100%	0.0	96%	669/697
S. aureus NCTC8325	1129	87%	0.0	96%	673/704
S. aureus VC40	1129	87%	0.0	96%	673/704

Table 2. *SAR1189* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1189	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1762	100%	0.0	100%	954/954
S. aureus RN4282	1679	100%	0.0	98%	939/954
S. aureus T1	1762	100%	0.0	100%	954/954
S. aureus COL	1679	100%	0.0	98%	939/954
S. aureus MSSA476	1679	100%	0.0	98%	939/954
S. aureus MW2	1679	100%	0.0	98%	939/954
S. aureus N315	1679	100%	0.0	98%	939/954
S. aureus Mu50	1679	100%	0.0	98%	939/954
S. aureus T0131	1679	100%	0.0	98%	939/954
S. aureus TW20	1679	100%	0.0	98%	939/954
S. aureus BMB9393	1679	100%	0.0	98%	939/954
S. aureus LGA251	448	58%	0.0	81%	453/557
S. aureus RF122	440	58%	0.0	81%	452/557
S. aureus ST398	436	58%	0.0	98%	939/954
S. aureus M013	1679	58%	0.0	81%	452/558
S. aureus M1	1679	100%	0.0	98%	939/954
S. aureus Mu3	1679	100%	0.0	98%	939/954
S. aureus str. Newman	1679	100%	0.0	98%	939/954
S. aureus NCTC8325	1679	100%	0.0	98%	939/954
S. aureus VC40	1679	100%	0.0	98%	939/954

Table 3. *SAR0439* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0439	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1430	100%	0.0	100%	774/774
S. aureus RN4282	1236	100%	0.0	97%	776/765
S. aureus T1	1430	100%	0.0	99%	774/772
S. aureus COL	773	100%	0.0	85%	659/777
S. aureus MSSA476	723	100%	0.0	84%	651/778
S. aureus MW2	723	100%	0.0	84%	651/778
S. aureus N315	1236	100%	0.0	96%	735/768
S. aureus Mu50	1236	100%	0.0	96%	735/768
S. aureus T0131	1029	100%	0.0	92%	678/738
S. aureus TW20	315	100%	0.0	76%	472/625
S. aureus BMB9393	1029	100%	0.0	92%	678/738
S. aureus LGA251	1099	99%	0.0	96%	656/686
S. aureus RF122	623	100%	0.0	83%	573/688
S. aureus ST398	649	99%	0.0	95%	547/644
S. aureus M013	529	100%	0.0	83%	499/603
S. aureus M1	747	100%	0.0	85%	630/741
S. aureus Mu3	1236	100%	0.0	96%	774/768
S. aureus str. Newman	1280	100%	0.0	97%	774/747
S. aureus NCTC8325	813	95%	0.0	87%	639/738
S. aureus VC40	813	95%	0.0	87%	679/738

Table 4. *SAR0442* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0442	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1424	100%	0.0	100%	771/771
S. aureus RN4282	1214	100%	0.0	93%	708/771
S. aureus T1	1424	100%	0.0	100%	771/771
S. aureus COL	1075	100%	0.0	92%	709/771
S. aureus MSSA476	1214	100%	0.0	95%	733/771
S. aureus MW2	1214	100%	0.0	95%	733/771
S. aureus N315	756	100%	0.0	89%	553/624
S. aureus Mu50	756	100%	0.0	89%	553/624
S. aureus T0131	1173	100%	0.0	94%	719/761
S. aureus TW20	313	80%	0.0	76%	482/634
S. aureus BMB9393	1173	100%	0.0	94%	719/761
S. aureus LGA251	669	87%	0.0	85%	576/680
S. aureus RF122	1173	100%	0.0	98%	668/684
S. aureus ST398	741	99%	0.0	88%	549/623
S. aureus M013	1042	100%	0.0	98%	584/594
S. aureus M1	1173	100%	0.0	94%	719/761
S. aureus Mu3	756	100%	0.0	89%	553/624
S. aureus str. Newman	1162	100%	0.0	94%	718/762
S. aureus NCTC8325	684	92%	0.0	84%	601/716
S. aureus VC40	684	92%	0.0	84%	601/716

Table 5. *SAR0206* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0206	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	2350	100%	0.0	100%	1272/1272
S. aureus RN4282	2272	100%	0.0	99%	1258/1272
S. aureus T1	2350	100%	0.0	100%	1272/1272
S. aureus COL	2272	100%	0.0	99%	1258/1272
S. aureus MSSA476	2272	100%	0.0	99%	1258/1272
S. aureus MW2	2266	100%	0.0	99%	1257/1272
S. aureus N315	2261	100%	0.0	99%	1256/1272
S. aureus Mu50	2350	100%	0.0	99%	1256/1272
S. aureus T0131	2338	100%	0.0	100%	1272/1272
S. aureus TW20	2350	100%	0.0	99%	1270/1272
S. aureus BMB9393	2350	100%	0.0	100%	1272/1272
S. aureus LGA251	2305	100%	0.0	99%	1264/1272
S. aureus RF122	2289	100%	0.0	99%	1261/1272
S. aureus ST398	2311	100%	0.0	99%	1265/1272
S. aureus M013	2278	100%	0.0	99%	1259/1272
S. aureus M1	2272	100%	0.0	99%	1258/1272
S. aureus Mu3	2261	100%	0.0	99%	1256/1272
S. aureus str. Newman	2272	100%	0.0	99%	1258/1272
S. aureus NCTC8325	2272	100%	0.0	99%	1258/1272
S. aureus VC40	2272	100%	0.0	99%	1258/1272

Table 6. *SAR0396* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0396	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1158	100%	0.0	100%	627/627
S. aureus RN4282	1125	100%	0.0	99%	621/627
S. aureus T1	1158	100%	0.0	100%	627/627
S. aureus COL	1125	100%	0.0	99%	621/627
S. aureus MSSA476	1120	100%	0.0	99%	620/627
S. aureus MW2	1120	100%	0.0	99%	621/627
S. aureus N315	1125	100%	0.0	99%	621/627
S. aureus Mu50	1125	100%	0.0	99%	621/627
S. aureus T0131	1125	100%	0.0	99%	621/627
S. aureus TW20	1125	100%	0.0	99%	621/627
S. aureus BMB9393	1125	100%	0.0	99%	621/627
S. aureus LGA251	1114	100%	0.0	99%	619/627
S. aureus RF122	1114	100%	0.0	99%	619/627
S. aureus ST398	1131	100%	0.0	99%	622/627
S. aureus M013	1125	100%	0.0	99%	621/627
S. aureus M1	1125	100%	0.0	99%	621/627
S. aureus Mu3	1125	100%	0.0	99%	621/627
S. aureus str. Newman	1125	100%	0.0	99%	621/627
S. aureus NCTC8325	1125	100%	0.0	99%	621/627
S. aureus VC40	1125	100%	0.0	99%	621/627

Table 7. *SAR0118* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0118	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1834	100%	0.0	100%	993/993
S. aureus RN4282	1801	100%	0.0	99%	987/993
S. aureus T1	1834	100%	0.0	100%	993/993
S. aureus COL	1801	100%	0.0	99%	987/993
S. aureus MSSA476	1784	100%	0.0	99%	984/993
S. aureus MW2	1784	100%	0.0	99%	987/993
S. aureus N315	1801	100%	0.0	89%	987/993
S. aureus Mu50	1801	100%	0.0	89%	987/993
S. aureus T0131	1834	100%	0.0	100%	993/993
S. aureus TW20	1834	100%	0.0	100%	993/993
S. aureus BMB9393	1834	100%	0.0	100%	993/993
S. aureus LGA251	1801	100%	0.0	99%	987/993
S. aureus RF122	1801	100%	0.0	99%	987/993
S. aureus ST398	1801	100%	0.0	99%	988/993
S. aureus M013	1807	100%	0.0	99%	988/993
S. aureus M1	1807	100%	0.0	99%	987/993
S. aureus Mu3	1801	100%	0.0	99%	987/993
S. aureus str. Newman	1801	100%	0.0	99%	987/993
S. aureus NCTC8325	1801	100%	0.0	99%	987/993
S. aureus VC40	1801	100%	0.0	99%	986/993

Table 8. *SAR0438* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0438	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1452	100%	0.0	100%	786/786
S. aureus RN4282	1452	100%	0.0	100%	786/786
S. aureus T1	1452	100%	0.0	100%	786/786
S. aureus COL	937	98%	0.0	92%	623/680
S. aureus MSSA476	545	80%	0.0	83%	506/609
S. aureus MW2	545	80%	0.0	83%	506/609
S. aureus N315	1181	100%	0.0	97%	682/703
S. aureus Mu50	1181	100%	0.0	97%	682/703
S. aureus T0131	931	98%	0.0	91%	622/680
S. aureus TW20	931	98%	0.0	91%	622/680
S. aureus BMB9393	931	98%	0.0	91%	622/680
S. aureus LGA251	1286	100%	0.0	96%	756/786
S. aureus RF122	1194	99%	0.0	94%	739/785
S. aureus ST398	1315	99%	0.0	97%	760/784
S. aureus M013	911	99%	0.0	88%	690/787
S. aureus M1	931	98%	0.0	91%	622/680
S. aureus Mu3	1181	100%	0.0	97%	682/703
S. aureus str. Newman	937	98%	0.0	92%	623/680
S. aureus NCTC8325	937	98%	0.0	92%	623/680
S. aureus VC40	937	98%	0.0	92%	623/680

Table 9. *SAR1034* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1034	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	2034	100%	0.0	100%	1101/1101
S. aureus RN4282	2017	100%	0.0	99%	1098/1101
S. aureus T1	2034	100%	0.0	100%	1101/1101
S. aureus COL	2017	100%	0.0	99%	1098/1101
S. aureus MSSA476	2012	100%	0.0	99%	1097/1101
S. aureus MW2	2012	100%	0.0	99%	1097/1101
S. aureus N315	2012	100%	0.0	99%	1097/1101
S. aureus Mu50	2012	100%	0.0	99%	1097/1101
S. aureus T0131	2017	100%	0.0	99%	1098/1101
S. aureus TW20	2017	100%	0.0	99%	1098/1101
S. aureus BMB9393	2017	100%	0.0	99%	1098/1101
S. aureus LGA251	2017	100%	0.0	99%	1098/1101
S. aureus RF122	2006	100%	0.0	99%	1096/1101
S. aureus ST398	2028	100%	0.0	99%	1100/1101
S. aureus M013	2017	100%	0.0	99%	1098/1101
S. aureus M1	2017	100%	0.0	99%	1098/1101
S. aureus Mu3	2017	100%	0.0	99%	1097/1101
S. aureus str. Newman	2017	100%	0.0	99%	1098/1101
S. aureus NCTC8325	2017	100%	0.0	99%	1098/1101
S. aureus VC40	2017	100%	0.0	99%	1097/1101

Table 10. *SAR1879* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1879	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1026	100%	0.0	100%	555/555
S. aureus RN4282	898	100%	0.0	96%	532/555
S. aureus T1	898	100%	0.0	100%	555/555
S. aureus COL	898	100%	0.0	96%	532/555
S. aureus MSSA476	915	100%	0.0	96%	535/555
S. aureus MW2	915	100%	0.0	96%	535/555
S. aureus N315	893	100%	0.0	96%	531/555
S. aureus Mu50	893	100%	0.0	96%	531/555
S. aureus T0131	898	100%	0.0	96%	531/555
S. aureus TW20	898	100%	0.0	96%	532/555
S. aureus BMB9393	893	100%	0.0	96%	532/555
S. aureus LGA251	891	100%	0.0	96%	531/555
S. aureus RF122	1003	100%	0.0	96%	547/555
S. aureus ST398	715	100%	0.0	96%	551/555
S. aureus M013	898	100%	0.0	96%	425/555
S. aureus M1	898	100%	0.0	96%	550/555
S. aureus Mu3	893	100%	0.0	96%	532/555
S. aureus str. Newman	893	100%	0.0	96%	531/555
S. aureus NCTC8325	898	100%	0.0	96%	532/555
S. aureus VC40	898	100%	0.0	96%	532/555

Table 11. *SAR0230* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0230	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	2726	100%	0.0	100%	1476/1476
S. aureus RN4282	2621	100%	0.0	99%	1457/1476
S. aureus T1	2726	100%	0.0	100%	1476/1476
S. aureus COL	2621	100%	0.0	99%	1457/1476
S. aureus MSSA476	2588	100%	0.0	98%	1451/1476
S. aureus MW2	2588	100%	0.0	98%	1451/1476
S. aureus N315	2599	100%	0.0	98%	1453/1476
S. aureus Mu50	2599	100%	0.0	98%	1453/1476
S. aureus T0131	2715	100%	0.0	99%	1474/1476
S. aureus TW20	2715	100%	0.0	99%	1476/1476
S. aureus BMB9393	2715	100%	0.0	99%	1474/1476
S. aureus LGA251	2588	100%	0.0	98%	1451/1476
S. aureus RF122	2615	100%	0.0	99%	1456/1476
S. aureus ST398	2649	100%	0.0	99%	1462/1476
S. aureus M013	2610	100%	0.0	99%	1455/1476
S. aureus M1	2621	100%	0.0	99%	1457/1476
S. aureus Mu3	2599	100%	0.0	98%	1453/1476
S. aureus str. Newman	2621	100%	0.0	99%	1457/1476
S. aureus NCTC8325	2621	100%	0.0	99%	1457/1476
S. aureus VC40	2621	100%	0.0	99%	1457/1476

Table 12. *SAR1066* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1066	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1158	100%	0.0	100%	627/627
S. aureus RN4282	1109	100%	0.0	99%	618/627
S. aureus T1	1158	100%	0.0	100%	627/627
S. aureus COL	1109	100%	0.0	99%	618/627
S. aureus MSSA476	1114	100%	0.0	99%	619/627
S. aureus MW2	1114	100%	0.0	99%	619/627
S. aureus N315	1103	100%	0.0	98%	617/627
S. aureus Mu50	1103	100%	0.0	98%	617/627
S. aureus T0131	1109	100%	0.0	99%	618/627
S. aureus TW20	1103	100%	0.0	98%	617/627
S. aureus BMB9393	1109	100%	0.0	99%	618/627
S. aureus LGA251	1048	100%	0.0	97%	607/627
S. aureus RF122	1042	100%	0.0	97%	606/627
S. aureus ST398	1131	100%	0.0	99%	622/627
S. aureus M013	1053	100%	0.0	97%	608/627
S. aureus M1	1109	100%	0.0	99%	618/627
S. aureus Mu3	1103	100%	0.0	98%	617/627
S. aureus str. Newman	1109	100%	0.0	99%	618/627
S. aureus NCTC8325	1109	100%	0.0	99%	618/627
S. aureus VC40	1109	100%	0.0	99%	618/627

Table 13. *SAR1881* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1881	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1158	100%	0.0	100%	627/627
S. aureus RN4282	959	100%	0.0	94%	591/627
S. aureus T1	1158	100%	0.0	100%	627/627
S. aureus COL	959	100%	0.0	94%	591/627
S. aureus MSSA476	987	100%	0.0	95%	596/627
S. aureus MW2	987	100%	0.0	95%	596/627
S. aureus N315	981	100%	0.0	95%	595/627
S. aureus Mu50	981	100%	0.0	95%	595/627
S. aureus T0131	953	100%	0.0	94%	590/627
S. aureus TW20	959	100%	0.0	94%	591/627
S. aureus BMB9393	959	100%	0.0	94%	591/627
S. aureus LGA251					
S. aureus RF122	1055	100%	0.0	97%	607/627
S. aureus ST398	1064	100%	0.0	97%	610/627
S. aureus M013	981	100%	0.0	98%	616/627
S. aureus M1	959	100%	0.0	94%	591/627
S. aureus Mu3	981	100%	0.0	95%	595/627
S. aureus str. Newman	959	100%	0.0	94%	591/627
S. aureus NCTC8325	959	100%	0.0	94%	591/627
S. aureus VC40	959	100%	0.0	94%	591/627

Table 14. *SAR0794* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0794	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1624	100%	0.0	100%	879/879
S. aureus RN4282	1613	100%	0.0	99%	877/879
S. aureus T1	1624	100%	0.0	100%	879/879
S. aureus COL	1613	100%	0.0	99%	877/879
S. aureus MSSA476	1613	100%	0.0	99%	877/879
S. aureus MW2	1613	100%	0.0	99%	877/879
S. aureus N315	1618	100%	0.0	99%	878/879
S. aureus Mu50	1618	100%	0.0	99%	878/879
S. aureus T0131	1596	100%	0.0	99%	878/879
S. aureus TW20	1613	100%	0.0	99%	877/879
S. aureus BMB9393	1613	100%	0.0	99%	877/879
S. aureus LGA251	1596	100%	0.0	99%	874/879
S. aureus RF122	1594	100%	0.0	99%	874/879
S. aureus ST398	1613	100%	0.0	99%	877/879
S. aureus M013	1613	100%	0.0	99%	877/879
S. aureus M1	1613	100%	0.0	99%	877/879
S. aureus Mu3	1618	100%	0.0	99%	878/879
S. aureus str. Newman	1613	100%	0.0	99%	877/879
S. aureus NCTC8325	1613	100%	0.0	99%	877/879
S. aureus VC40	1613	100%	0.0	99%	877/879

Table 15. *SAR2104* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2104	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252		100%	0.0	100%	
S. aureus RN4282		100%	0.0	93%	
S. aureus T1		100%	0.0	100%	
S. aureus COL					
S. aureus MSSA476					
S. aureus MW2					
S. aureus N315					
S. aureus Mu50		100%	0.0	89%	
S. aureus T0131					
S. aureus TW20		100%	0.0	76%	
S. aureus BMB9393					
S. aureus LGA251					
S. aureus RF122					
S. aureus ST398		100%	0.0	88%	
S. aureus M013					
S. aureus M1					
S. aureus Mu3		100%	0.0	89%	
S. aureus str. Newman		100%	0.0	94%	
S. aureus NCTC8325					
S. aureus VC40					

Table 16. *SAR0872* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0872	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1519	100%	0.0	100%	822/822
S. aureus RN4282	1519	100%	0.0	99%	813/822
S. aureus T1	1519	100%	0.0	100%	822/822
S. aureus COL	1519	100%	0.0	99%	813/822
S. aureus MSSA476	1463	100%	0.0	99%	812/822
S. aureus MW2	1469	100%	0.0	99%	813/822
S. aureus N315	1488	100%	0.0	99%	815/822
S. aureus Mu50	1488	100%	0.0	99%	815/822
S. aureus T0131	1469	100%	0.0	99%	813/822
S. aureus TW20	1469	100%	0.0	99%	813/822
S. aureus BMB9393	1469	100%	0.0	99%	813/822
S. aureus LGA251	1452	100%	0.0	99%	810/822
S. aureus RF122	1458	100%	0.0	99%	811/822
S. aureus ST398	1519	100%	0.0	100%	822/822
S. aureus M013	1485	100%	0.0	99%	816/822
S. aureus M1	1469	100%	0.0	99%	813/822
S. aureus Mu3	1480	100%	0.0	99%	815/822
S. aureus str. Newman	1463	100%	0.0	99%	812/822
S. aureus NCTC8325	1469	100%	0.0	99%	813/822
S. aureus VC40	1469	100%	0.0	99%	813/822

Table 17. *SAR0790* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0790	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1901	100%	0.0	100%	1029/1029
S. aureus RN4282	1557	100%	0.0	94%	967/1029
S. aureus T1	1901	100%	0.0	100%	1029/1029
S. aureus COL	1550	100%	0.0	94%	967/1029
S. aureus MSSA476	1563	100%	0.0	99%	968/1029
S. aureus MW2	1563	100%	0.0	99%	968/1029
S. aureus N315	1563	100%	0.0	99%	968/1029
S. aureus Mu50	1563	100%	0.0	99%	968/1029
S. aureus T0131	1557	100%	0.0	94%	967/1029
S. aureus TW20	1557	100%	0.0	94%	967/1029
S. aureus BMB9393	1557	100%	0.0	94%	967/1029
S. aureus LGA251	1507	100%	0.0	93%	958/1029
S. aureus RF122	1360	100%	0.0	91%	936/1029
S. aureus ST398	1840	100%	0.0	99%	968/1029
S. aureus M013	1602	100%	0.0	95%	975/1029
S. aureus M1	1552	100%	0.0	94%	966/1029
S. aureus Mu3	1563	100%	0.0	99%	968/1029
S. aureus str. Newman	1557	100%	0.0	94%	967/1029
S. aureus NCTC8325	1557	100%	0.0	94%	967/1029
S. aureus VC40	1557	100%	0.0	94%	967/1029

Table 18. *SAR1106* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1106	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1624	100%	0.0	100%	879/789
S. aureus RN4282	1613	100%	0.0	99%	877/789
S. aureus T1	1624	100%	0.0	100%	879/789
S. aureus COL	1613	100%	0.0	99%	877/789
S. aureus MSSA476	1596	100%	0.0	99%	874/789
S. aureus MW2	1596	100%	0.0	99%	874/789
S. aureus N315	1613	100%	0.0	99%	877/789
S. aureus Mu50	1613	100%	0.0	99%	877/789
S. aureus T0131	1607	100%	0.0	99%	876/789
S. aureus TW20	1613	100%	0.0	99%	877/789
S. aureus BMB9393	1613	100%	0.0	99%	877/789
S. aureus LGA251	1613	100%	0.0	99%	877/789
S. aureus RF122	1591	100%	0.0	99%	873/789
S. aureus ST398	1602	100%	0.0	99%	875/789
S. aureus M013	1602	100%	0.0	99%	875/789
S. aureus M1	1613	100%	0.0	99%	877/789
S. aureus Mu3	1613	100%	0.0	99%	877/789
S. aureus str. Newman	1613	100%	0.0	99%	877/789
S. aureus NCTC8325	1613	100%	0.0	99%	877/789
S. aureus VC40	1613	100%	0.0	99%	877/789

Table 19. *SAR0953* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0953	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	3059	100%	0.0	100%	1657/1657
S. aureus RN4282	2809	100%	0.0	97%	1612/1657
S. aureus T1	3059	100%	0.0	100%	1657/1657
S. aureus COL	2809	100%	0.0	97%	1612/1657
S. aureus MSSA476	2826	100%	0.0	97%	1615/1657
S. aureus MW2	2820	100%	0.0	97%	1614/1657
S. aureus N315	2804	100%	0.0	97%	1611/1657
S. aureus Mu50	2804	100%	0.0	97%	1611/1657
S. aureus T0131	2809	100%	0.0	97%	1612/1657
S. aureus TW20	2809	100%	0.0	97%	1612/1657
S. aureus BMB9393	2809	100%	0.0	97%	1612/1657
S. aureus LGA251	2865	100%	0.0	98%	1621/1656
S. aureus RF122	2854	100%	0.0	98%	1640/1656
S. aureus ST398	2970	100%	0.0	99%	1618/1657
S. aureus M013	2809	100%	0.0	98%	1618/1656
S. aureus M1	2809	100%	0.0	97%	1612/1657
S. aureus Mu3	2804	100%	0.0	97%	1611/1657
S. aureus str. Newman	2809	100%	0.0	97%	1612/1657
S. aureus NCTC8325	2809	100%	0.0	97%	1612/1657
S. aureus VC40	2809	100%	0.0	97%	1612/1657

Table 20. *SAR1011* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1011	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1773	100%	0.0	100%	960/960
S. aureus RN4282	1585	100%	0.0	96%	926/960
S. aureus T1	1773	100%	0.0	100%	960/960
S. aureus COL	1585	100%	0.0	96%	926/960
S. aureus MSSA476	1580	100%	0.0	96%	925/960
S. aureus MW2	1580	100%	0.0	96%	925/960
S. aureus N315	1602	100%	0.0	97%	929/960
S. aureus Mu50	1602	100%	0.0	97%	929/960
S. aureus T0131	1585	100%	0.0	96%	926/960
S. aureus TW20	1585	100%	0.0	96%	926/960
S. aureus BMB9393	1585	100%	0.0	96%	926/960
S. aureus LGA251	1640	100%	0.0	98%	936/960
S. aureus RF122	1554	100%	0.0	96%	921/960
S. aureus ST398	1657	100%	0.0	98%	939/960
S. aureus M013	1568	100%	0.0	96%	923/960
S. aureus M1	1585	100%	0.0	96%	926/960
S. aureus Mu3	1602	100%	0.0	98%	929/960
S. aureus str. Newman	1585	100%	0.0	96%	926/960
S. aureus NCTC8325	1585	100%	0.0	96%	926/960
S. aureus VC40	1585	100%	0.0	96%	926/960

Table 21. *SAR1995* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1995	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	2217	100%	0.0	100%	1200/1200
S. aureus RN4282	2178	100%	0.0	99%	1193/1200
S. aureus T1	2217	100%	0.0	100%	1200/1200
S. aureus COL	2178	100%	0.0	99%	1193/1200
S. aureus MSSA476	2178	100%	0.0	99%	1193/1200
S. aureus MW2	2178	100%	0.0	99%	1193/1200
S. aureus N315	2178	100%	0.0	99%	1193/1200
S. aureus Mu50	2178	100%	0.0	99%	1193/1200
S. aureus T0131	2178	100%	0.0	99%	1193/1200
S. aureus TW20	2178	100%	0.0	99%	1193/1200
S. aureus BMB9393	2178	100%	0.0	99%	1193/1200
S. aureus LGA251	2172	100%	0.0	99%	1192/1200
S. aureus RF122	2178	100%	0.0	99%	1193/1200
S. aureus ST398	2206	100%	0.0	99%	1198/1200
S. aureus M013	2167	100%	0.0	99%	1191/1200
S. aureus M1	2178	100%	0.0	99%	1193/1200
S. aureus Mu3	2178	100%	0.0	99%	1193/1200
S. aureus str. Newman	2178	100%	0.0	99%	1193/1200
S. aureus NCTC8325	2178	100%	0.0	99%	1193/1200
S. aureus VC40	2178	100%	0.0	99%	1193/1200

Table 22. *SAR2179* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2179	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1613	100%	0.0	100%	873/873
S. aureus RN4282	1568	100%	0.0	99%	865/873
S. aureus T1	1613	100%	0.0	100%	873/873
S. aureus COL	1568	100%	0.0	99%	865/873
S. aureus MSSA476	1580	100%	0.0	99%	867/873
S. aureus MW2	1580	100%	0.0	99%	867/873
S. aureus N315	1568	100%	0.0	99%	865/873
S. aureus Mu50	1568	100%	0.0	99%	865/873
S. aureus T0131	1555	100%	0.0	99%	863/873
S. aureus TW20	1568	100%	0.0	99%	865/873
S. aureus BMB9393	1568	100%	0.0	99%	863/873
S. aureus LGA251	1563	100%	0.0	99%	865/873
S. aureus RF122	1574	100%	0.0	99%	865/873
S. aureus ST398	1574	100%	0.0	99%	866/873
S. aureus M013	1563	100%	0.0	99%	866/873
S. aureus M1	1563	100%	0.0	99%	864/873
S. aureus Mu3	1568	100%	0.0	99%	865/873
S. aureus str. Newman	1568	100%	0.0	99%	865/873
S. aureus NCTC8325	1568	100%	0.0	99%	865/873
S. aureus VC40	1568	100%	0.0	99%	865/873

Table 23. *SAR1494* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1494	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1679	100%	0.0	100%	909/909
S. aureus RN4282	1392	100%	0.0	94%	851/909
S. aureus T1	1679	100%	0.0	100%	909/909
S. aureus COL	1373	100%	0.0	94%	851/905
S. aureus MSSA476	929	70%	0.0	96%	549/572
S. aureus MW2	924	70%	0.0	96%	548/572
S. aureus N315	918	68%	0.0	96%	547/572
S. aureus Mu50	918	68%	0.0	96%	852/905
S. aureus T0131	1378	100%	0.0	94%	850/905
S. aureus TW20	1367	100%	0.0	94%	851/905
S. aureus BMB9393	1373	99%	0.0	94%	851/905
S. aureus LGA251	1469	91%	0.0	99%	817/828
S. aureus RF122	1395	100%	0.0	94%	858/909
S. aureus ST398	732	71%	0.0	94%	851/905
S. aureus M013	941	68%	0.0	96%	552/573
S. aureus M1	1373	99%	0.0	94%	851/905
S. aureus Mu3	918	68%	0.0	96%	547/572
S. aureus str. Newman	1373	99%	0.0	94%	851/909
S. aureus NCTC8325	1373	99%	0.0	94%	851/909
S. aureus VC40	1373	99%	0.0	94%	851/909

Table 24. *SAR1608* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1608	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1075	100%	0.0	100%	582/582
S. aureus RN4282	1059	100%	0.0	99%	579/582
S. aureus T1	1075	100%	0.0	100%	582/582
S. aureus COL	1059	100%	0.0	99%	579/582
S. aureus MSSA476	1044	100%	0.0	99%	579/582
S. aureus MW2	1044	100%	0.0	99%	579/582
S. aureus N315	1059	100%	0.0	99%	579/582
S. aureus Mu50	1059	100%	0.0	99%	579/582
S. aureus T0131	1059	100%	0.0	99%	579/582
S. aureus TW20	1059	100%	0.0	99%	579/582
S. aureus BMB9393	1059	100%	0.0	99%	579/582
S. aureus LGA251	1059	100%	0.0	99%	579/582
S. aureus RF122	1048	100%	0.0	99%	577/582
S. aureus ST398	1053	100%	0.0	99%	578/582
S. aureus M013	1042	100%	0.0	99%	576/582
S. aureus M1	1059	100%	0.0	99%	579/582
S. aureus Mu3	1059	100%	0.0	99%	579/582
S. aureus str. Newman	1059	100%	0.0	99%	579/582
S. aureus NCTC8325	1059	100%	0.0	99%	579/582
S. aureus VC40	1059	100%	0.0	99%	579/582

Table 25. *SAR2368* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2368	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1674	100%	0.0	100%	909/909
S. aureus RN4282	1629	100%	0.0	99%	900/909
S. aureus T1	1674	100%	0.0	100%	909/909
S. aureus COL	1629	100%	0.0	99%	900/909
S. aureus MSSA476	1629	100%	0.0	99%	900/909
S. aureus MW2	1629	100%	0.0	99%	900/909
S. aureus N315	1624	100%	0.0	99%	899/909
S. aureus Mu50	1624	100%	0.0	99%	899/909
S. aureus T0131	1629	100%	0.0	99%	900/909
S. aureus TW20	1629	100%	0.0	99%	900/909
S. aureus BMB9393	1629	100%	0.0	99%	900/909
S. aureus LGA251	1629	100%	0.0	99%	900/909
S. aureus RF122	1618	100%	0.0	99%	898/909
S. aureus ST398	1646	100%	0.0	99%	903/909
S. aureus M013	1629	100%	0.0	99%	900/909
S. aureus M1	1629	100%	0.0	99%	900/909
S. aureus Mu3	1624	100%	0.0	99%	899/909
S. aureus str. Newman	1629	100%	0.0	99%	900/909
S. aureus NCTC8325	1629	100%	0.0	99%	900/909
S. aureus VC40	1629	100%	0.0	99%	900/909

Table 26. *SAR2457* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2457	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1164	100%	0.0	100%	630/630
S. aureus RN4282	931	100%	0.0	93%	588/630
S. aureus T1	1164	100%	0.0	100%	630/630
S. aureus COL	931	100%	0.0	93%	588/630
S. aureus MSSA476	883	100%	0.0	92%	582/630
S. aureus MW2	883	100%	0.0	92%	582/630
S. aureus N315	900	100%	0.0	93%	585/630
S. aureus Mu50	900	100%	0.0	93%	585/630
S. aureus T0131	931	100%	0.0	93%	588/630
S. aureus TW20	931	100%	0.0	93%	588/630
S. aureus BMB9393	931	100%	0.0	93%	588/630
S. aureus LGA251	942	100%	0.0	94%	590/630
S. aureus RF122	931	100%	0.0	93%	585/630
S. aureus ST398	1153	100%	0.0	99%	628/630
S. aureus M013	948	100%	0.0	93%	588/630
S. aureus M1	931	100%	0.0	93%	588/630
S. aureus Mu3	900	100%	0.0	93%	585/630
S. aureus str. Newman	931	100%	0.0	93%	588/630
S. aureus NCTC8325	931	100%	0.0	93%	588/630
S. aureus VC40	931	100%	0.0	93%	588/630

Table 27. *SAR2268* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2268	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1818	100%	0.0	100%	984/984
S. aureus RN4282	1657	100%	0.0	97%	955/984
S. aureus T1	1818	100%	0.0	100%	984/984
S. aureus COL	1657	100%	0.0	97%	955/984
S. aureus MSSA476	1635	100%	0.0	97%	951/984
S. aureus MW2	1635	100%	0.0	97%	951/984
S. aureus N315	1663	100%	0.0	97%	959/984
S. aureus Mu50	1663	100%	0.0	97%	956/984
S. aureus T0131	1657	100%	0.0	97%	955/984
S. aureus TW20	1657	100%	0.0	97%	955/984
S. aureus BMB9393	1657	100%	0.0	97%	955/984
S. aureus LGA251	1696	100%	0.0	99%	962/984
S. aureus RF122	1696	100%	0.0	99%	962/984
S. aureus ST398	1779	100%	0.0	99%	977/984
S. aureus M013	1690	100%	0.0	97%	961/984
S. aureus M1	1657	100%	0.0	97%	955/984
S. aureus Mu3	1663	100%	0.0	97%	956/984
S. aureus str. Newman	1657	100%	0.0	97%	955/984
S. aureus NCTC8325	1657	100%	0.0	97%	955/984
S. aureus VC40	1657	100%	0.0	97%	955/984

Table 28. *SAR2363* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2363	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1447	100%	0.0	100%	783/783
S. aureus RN4282	1225	100%	0.0	97%	743/783
S. aureus T1	1447	100%	0.0	100%	783/783
S. aureus COL	1225	100%	0.0	95%	743/783
S. aureus MSSA476	1286	100%	0.0	96%	754/783
S. aureus MW2	1286	100%	0.0	96%	754/783
S. aureus N315	1286	100%	0.0	96%	754/783
S. aureus Mu50	1286	100%	0.0	96%	754/783
S. aureus T0131	1225	100%	0.0	95%	743/783
S. aureus TW20	1219	100%	0.0	95%	742/783
S. aureus BMB9393	1225	100%	0.0	95%	743/783
S. aureus LGA251	1236	100%	0.0	95%	745/783
S. aureus RF122	1236	100%	0.0	95%	746/783
S. aureus ST398	1424	100%	0.0	99%	779/783
S. aureus M013	1225	100%	0.0	95%	743/783
S. aureus M1	1225	100%	0.0	95%	743/783
S. aureus Mu3	1286	100%	0.0	96%	754/783
S. aureus str. Newman	1225	100%	0.0	95%	743/783
S. aureus NCTC8325	1225	100%	0.0	95%	743/783
S. aureus VC40	1218	100%	0.0	95%	742/783

Table 29. *SAR2500* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2500	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	671	100%	0.0	100%	363/363
S. aureus RN4282	671	100%	0.0	100%	363/363
S. aureus T1	671	100%	0.0	100%	363/363
S. aureus COL	671	100%	0.0	100%	363/363
S. aureus MSSA476	671	100%	0.0	100%	363/363
S. aureus MW2	671	100%	0.0	100%	363/363
S. aureus N315	671	100%	0.0	100%	363/363
S. aureus Mu50	671	100%	0.0	100%	363/363
S. aureus T0131	671	100%	0.0	100%	363/363
S. aureus TW20	671	100%	0.0	100%	363/363
S. aureus BMB9393	671	100%	0.0	100%	363/363
S. aureus LGA251	671	100%	0.0	100%	363/363
S. aureus RF122	671	100%	0.0	100%	363/363
S. aureus ST398	660	100%	0.0	99%	362/363
S. aureus M013	660	100%	0.0	99%	362/363
S. aureus M1	671	100%	0.0	100%	363/363
S. aureus Mu3	671	100%	0.0	100%	363/363
S. aureus str. Newman	671	100%	0.0	100%	363/363
S. aureus NCTC8325	671	100%	0.0	100%	363/363
S. aureus VC40	671	100%	0.0	100%	363/363

Table 30. *SAR2536* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2536	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1740	100%	0.0	100%	942/942
S. aureus RN4282	1663	100%	0.0	99%	932/942
S. aureus T1	1740	100%	0.0	100%	942/942
S. aureus COL	1663	100%	0.0	99%	932/942
S. aureus MSSA476	1663	100%	0.0	99%	932/942
S. aureus MW2	1685	100%	0.0	99%	932/942
S. aureus N315	1685	100%	0.0	99%	932/942
S. aureus Mu50	1685	100%	0.0	99%	932/942
S. aureus T0131	1663	100%	0.0	99%	932/942
S. aureus TW20	1663	100%	0.0	99%	932/942
S. aureus BMB9393	1663	100%	0.0	99%	932/942
S. aureus LGA251	1657	100%	0.0	98%	927/942
S. aureus RF122	1646	100%	0.0	98%	927/942
S. aureus ST398	1657	100%	0.0	98%	927/942
S. aureus M013	1640	100%	0.0	98%	927/942
S. aureus M1	1663	100%	0.0	99%	932/942
S. aureus Mu3	1685	100%	0.0	99%	932/942
S. aureus str. Newman	1663	100%	0.0	99%	932/942
S. aureus NCTC8325	1663	100%	0.0	99%	932/942
S. aureus VC40	1663	100%	0.0	99%	932/942

Table 31. *SAR2504* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2504	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1441	100%	0.0	100%	780/780
S. aureus RN4282	1291	100%	0.0	97%	753/780
S. aureus T1	1441	100%	0.0	100%	780/780
S. aureus COL	1291	100%	0.0	97%	753/780
S. aureus MSSA476	1291	100%	0.0	97%	753/780
S. aureus MW2	1291	100%	0.0	97%	753/780
S. aureus N315	1303	100%	0.0	97%	755/780
S. aureus Mu50	1303	100%	0.0	97%	755/780
S. aureus T0131	1291	100%	0.0	97%	753/780
S. aureus TW20	1291	100%	0.0	97%	753/780
S. aureus BMB9393	1291	100%	0.0	97%	753/780
S. aureus LGA251	1314	100%	0.0	97%	757/780
S. aureus RF122	1286	100%	0.0	96%	752/780
S. aureus ST398	1408	100%	0.0	99%	774/780
S. aureus M013	1291	100%	0.0	97%	753/780
S. aureus M1	1291	100%	0.0	97%	753/780
S. aureus Mu3	1303	100%	0.0	97%	753/780
S. aureus str. Newman	1291	100%	0.0	97%	753/780
S. aureus NCTC8325	1291	100%	0.0	97%	753/780
S. aureus VC40	1291	100%	0.0	97%	753/780

Table 32. *SAR2499* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2499	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1109	100%	0.0	100%	600/600
S. aureus RN4282	1075	100%	0.0	99%	594/600
S. aureus T1	1109	100%	0.0	100%	600/600
S. aureus COL	1075	100%	0.0	99%	594/600
S. aureus MSSA476	1070	100%	0.0	99%	593/600
S. aureus MW2	1070	100%	0.0	99%	593/600
S. aureus N315	1070	100%	0.0	99%	593/600
S. aureus Mu50	1070	100%	0.0	99%	593/600
S. aureus T0131	1075	100%	0.0	99%	594/600
S. aureus TW20	1075	100%	0.0	99%	594/600
S. aureus BMB9393	1075	100%	0.0	99%	594/600
S. aureus LGA251	1075	100%	0.0	99%	594/600
S. aureus RF122	1092	100%	0.0	99%	597/600
S. aureus ST398	1064	100%	0.0	99%	592/600
S. aureus M013	1075	100%	0.0	99%	594/600
S. aureus M1	1075	100%	0.0	99%	594/600
S. aureus Mu3	1070	100%	0.0	99%	593/600
S. aureus str. Newman	1075	100%	0.0	99%	594/600
S. aureus NCTC8325	1075	100%	0.0	99%	594/600
S. aureus VC40	1075	100%	0.0	99%	594/600

Table 33. *SAR2546* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2546	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	843	100%	0.0	100%	456/456
S. aureus RN4282	826	100%	0.0	99%	453/456
S. aureus T1	843	100%	0.0	100%	456/456
S. aureus COL	826	100%	0.0	99%	453/456
S. aureus MSSA476	837	100%	0.0	99%	455/456
S. aureus MW2	837	100%	0.0	99%	455/456
S. aureus N315	821	100%	0.0	99%	452/456
S. aureus Mu50	821	100%	0.0	99%	452/456
S. aureus T0131	286	100%	0.0	99%	453/456
S. aureus TW20	821	100%	0.0	99%	452/456
S. aureus BMB9393	826	100%	0.0	99%	453/456
S. aureus LGA251	832	100%	0.0	99%	454/456
S. aureus RF122	832	100%	0.0	99%	454/456
S. aureus ST398	837	100%	0.0	99%	455/456
S. aureus M013	832	100%	0.0	99%	454/456
S. aureus M1	826	100%	0.0	99%	453/456
S. aureus Mu3	821	100%	0.0	99%	452/456
S. aureus str. Newman	826	100%	0.0	99%	453/456
S. aureus NCTC8325	826	100%	0.0	99%	453/456
S. aureus VC40	826	100%	0.0	99%	453/456

Table 34. *SAR2554* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2554	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	2953	100%	0.0	100%	1599/1599
S. aureus RN4282	2854	100%	0.0	99%	1580/1599
S. aureus T1	2953	100%	0.0	100%	1599/1599
S. aureus COL	2854	100%	0.0	99%	1575/1599
S. aureus MSSA476	2820	100%	0.0	98%	1575/1599
S. aureus MW2	2820	100%	0.0	98%	1575/1599
S. aureus N315	2798	100%	0.0	98%	1571/1599
S. aureus Mu50	2798	100%	0.0	98%	1571/1599
S. aureus T0131	2854	100%	0.0	99%	1581/1599
S. aureus TW20	2854	100%	0.0	99%	1581/1599
S. aureus BMB9393	2854	100%	0.0	99%	1581/1599
S. aureus LGA251	2870	100%	0.0	99%	1584/1599
S. aureus RF122	2843	100%	0.0	99%	1579/1599
S. aureus ST398	2915	100%	0.0	99%	1592/1599
S. aureus M013	2865	100%	0.0	99%	1583/1599
S. aureus M1	2854	100%	0.0	99%	1581/1599
S. aureus Mu3	2798	100%	0.0	98%	1571/1599
S. aureus str. Newman	2854	100%	0.0	99%	1580/1599
S. aureus NCTC8325	2845	100%	0.0	99%	1580/1599
S. aureus VC40	2845	100%	0.0	99%	1580/1599

Table 35. *SAR0463* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0463	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1557	100%	0.0	100%	843/843
S. aureus RN4282	1458	100%	0.0	98%	825/843
S. aureus T1	1552	100%	0.0	99%	842/843
S. aureus COL	1458	100%	0.0	98%	825/843
S. aureus MSSA476	1435	100%	0.0	98%	821/843
S. aureus MW2	1441	100%	0.0	98%	822/843
S. aureus N315	1452	100%	0.0	98%	824/843
S. aureus Mu50	1452	100%	0.0	98%	824/843
S. aureus T0131	1452	100%	0.0	98%	824/843
S. aureus TW20	1458	100%	0.0	98%	825/843
S. aureus BMB9393	1458	100%	0.0	98%	825/843
S. aureus LGA251	1458	100%	0.0	98%	830/843
S. aureus RF122	1458	100%	0.0	98%	830/843
S. aureus ST398	1474	100%	0.0	98%	828/843
S. aureus M013	1480	100%	0.0	98%	829/843
S. aureus M1	1458	100%	0.0	98%	825/843
S. aureus Mu3	1452	100%	0.0	98%	824/843
S. aureus str. Newman	1458	100%	0.0	98%	825/843
S. aureus NCTC8325	1458	100%	0.0	98%	825/843
S. aureus VC40	1458	100%	0.0	98%	825/843

Table 36. *SAR0444* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0444	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1469	100%	0.0	100%	795/795
S. aureus RN4282	1208	100%	0.0	94%	749/795
S. aureus T1	1469	100%	0.0	100%	795/795
S. aureus COL	1190	100%	0.0	94%	742/790
S. aureus MSSA476	802	100%	0.0	88%	613/700
S. aureus MW2	802	100%	0.0	88%	613/700
S. aureus N315	1164	100%	0.0	94%	735/795
S. aureus Mu50	1164	100%	0.0	94%	735/795
S. aureus T0131	1190	100%	0.0	94%	742/790
S. aureus TW20	1026	100%	0.0	93%	651/698
S. aureus BMB9393	734	100%	0.0	86%	592/686
S. aureus LGA251	1120	100%	0.0	93%	723/780
S. aureus RF122	710	100%	0.0	95%	611/720
S. aureus ST398	1397	100%	0.0	98%	783/795
S. aureus M013	771	100%	0.0	86%	623/723
S. aureus M1	1190	100%	0.0	94%	742/790
S. aureus Mu3	1164	100%	0.0	94%	735/786
S. aureus str. Newman	1190	100%	0.0	94%	742/790
S. aureus NCTC8325	1190	100%	0.0	94%	742/790
S. aureus VC40	1190	100%	0.0	94%	742/790

Table 37. *SAR0618* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0618	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1640	100%	0.0	100%	888/888
S. aureus RN4282	1602	100%	0.0	99%	881/888
S. aureus T1	1640	100%	0.0	100%	888/888
S. aureus COL	1596	100%	0.0	99%	880/888
S. aureus MSSA476	1585	100%	0.0	99%	878/888
S. aureus MW2	1585	100%	0.0	99%	878/888
S. aureus N315	1591	100%	0.0	99%	879/888
S. aureus Mu50	1591	100%	0.0	99%	879/888
S. aureus T0131	1596	100%	0.0	99%	880/888
S. aureus TW20	1596	100%	0.0	99%	880/888
S. aureus BMB9393	1596	100%	0.0	99%	880/888
S. aureus LGA251	1607	100%	0.0	99%	882/888
S. aureus RF122	1607	100%	0.0	99%	882/888
S. aureus ST398	1591	100%	0.0	99%	879/888
S. aureus M013	1607	100%	0.0	99%	882/888
S. aureus M1	1591	100%	0.0	99%	879/888
S. aureus Mu3	1591	100%	0.0	99%	879/888
S. aureus str. Newman	1596	100%	0.0	99%	880/888
S. aureus NCTC8325	1596	100%	0.0	99%	880/888
S. aureus VC40	1596	100%	0.0	99%	880/888

Table 38. *SAR0641* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0641	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1718	100%	0.0	100%	930/930
S. aureus RN4282	1696	100%	0.0	99%	926/930
S. aureus T1	1718	100%	0.0	100%	930/930
S. aureus COL	1696	100%	0.0	99%	926/930
S. aureus MSSA476	1696	100%	0.0	99%	926/930
S. aureus MW2	1696	100%	0.0	99%	926/930
S. aureus N315	1690	100%	0.0	99%	925/930
S. aureus Mu50	1690	100%	0.0	99%	925/930
S. aureus T0131	1696	100%	0.0	99%	926/930
S. aureus TW20	1696	100%	0.0	99%	926/930
S. aureus BMB9393	1696	100%	0.0	99%	926/930
S. aureus LGA251	1690	100%	0.0	99%	925/930
S. aureus RF122	1701	100%	0.0	99%	927/930
S. aureus ST398	1701	100%	0.0	99%	927/930
S. aureus M013	1701	100%	0.0	99%	927/930
S. aureus M1	1696	100%	0.0	99%	926/930
S. aureus Mu3	1690	100%	0.0	99%	925/930
S. aureus str. Newman	1696	100%	0.0	99%	926/930
S. aureus NCTC8325	1659	100%	0.0	99%	927/930
S. aureus VC40	1659	100%	0.0	99%	927/930

Table 39. *SAR0730* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0730	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	721	100%	0.0	100%	573/573
S. aureus RN4282	693	100%	0.0	99%	566/573
S. aureus T1	721	100%	0.0	100%	573/573
S. aureus COL	660	100%	0.0	97%	566/573
S. aureus MSSA476	660	100%	0.0	97%	567/573
S. aureus MW2	654	100%	0.0	97%	567/573
S. aureus N315	665	100%	0.0	97%	566/573
S. aureus Mu50	665	100%	0.0	97%	566/573
S. aureus T0131	660	100%	0.0	97%	566/573
S. aureus TW20	660	100%	0.0	97%	566/573
S. aureus BMB9393	660	100%	0.0	97%	566/573
S. aureus LGA251	521	100%	0.0	91%	567/573
S. aureus RF122	688	100%	0.0	98%	567/573
S. aureus ST398	499	100%	0.0	90%	530/573
S. aureus M013	593	100%	0.0	94%	540/573
S. aureus M1	660	100%	0.0	97%	566/573
S. aureus Mu3	665	100%	0.0	97%	566/573
S. aureus str. Newman	660	100%	0.0	97%	566/573
S. aureus NCTC8325	660	100%	0.0	97%	566/573
S. aureus VC40	660	100%	0.0	97%	566/573

Table 40. *SAR0390* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0390	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1059	100%	0.0	100%	573/573
S. aureus RN4282	1020	100%	0.0	99%	566/573
S. aureus T1	1059	100%	0.0	100%	573/573
S. aureus COL	1020	100%	0.0	99%	566/573
S. aureus MSSA476	1026	100%	0.0	99%	567/573
S. aureus MW2	1020	100%	0.0	99%	567/573
S. aureus N315	1020	100%	0.0	99%	566/573
S. aureus Mu50	1020	100%	0.0	99%	566/573
S. aureus T0131	1020	100%	0.0	99%	566/573
S. aureus TW20	1020	100%	0.0	99%	566/573
S. aureus BMB9393	1020	100%	0.0	99%	566/573
S. aureus LGA251	1026	100%	0.0	99%	567/573
S. aureus RF122	1014	100%	0.0	99%	566/573
S. aureus ST398	1020	100%	0.0	99%	566/573
S. aureus M013	1020	100%	0.0	99%	566/573
S. aureus M1	1020	100%	0.0	99%	566/573
S. aureus Mu3	1020	100%	0.0	99%	566/573
S. aureus str. Newman	1020	100%	0.0	99%	566/573
S. aureus NCTC8325	1020	100%	0.0	99%	566/573
S. aureus VC40	1020	100%	0.0	99%	566/573

Table 41. *SAR0340* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0340	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1580	100%	0.0	100%	855/855
S. aureus RN4282	1546	100%	0.0	99%	849/855
S. aureus T1	1574	100%	0.0	99%	854/855
S. aureus COL	1552	100%	0.0	99%	850/855
S. aureus MSSA476	1535	100%	0.0	99%	847/855
S. aureus MW2	1535	100%	0.0	99%	847/855
S. aureus N315	1557	100%	0.0	99%	851/855
S. aureus Mu50	1557	100%	0.0	99%	851/855
S. aureus T0131	1552	100%	0.0	99%	850/855
S. aureus TW20	1544	100%	0.0	99%	849/855
S. aureus BMB9393	1552	100%	0.0	99%	850/855
S. aureus LGA251	1541	100%	0.0	99%	848/855
S. aureus RF122	1541	100%	0.0	99%	848/855
S. aureus ST398	1530	100%	0.0	99%	846/855
S. aureus M013	1546	100%	0.0	99%	849/855
S. aureus M1	1527	100%	0.0	99%	850/855
S. aureus Mu3	1527	100%	0.0	99%	851/855
S. aureus str. Newman	1522	100%	0.0	99%	850/855
S. aureus NCTC8325	1522	100%	0.0	99%	850/855
S. aureus VC40	1522	100%	0.0	99%	850/855

Table 42. *SAR0216* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0216	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1790	100%	0.0	100%	969/969
S. aureus RN4282	1690	100%	0.0	98%	951/969
S. aureus T1	1790	100%	0.0	100%	969/969
S. aureus COL	1707	100%	0.0	98%	954/969
S. aureus MSSA476	1701	100%	0.0	98%	953/969
S. aureus MW2	1701	100%	0.0	98%	953/969
S. aureus N315	1696	100%	0.0	98%	952/969
S. aureus Mu50	1696	100%	0.0	98%	952/969
S. aureus T0131	1784	100%	0.0	99%	968/969
S. aureus TW20	1790	100%	0.0	100%	969/969
S. aureus BMB9393	1790	100%	0.0	100%	969/969
S. aureus LGA251	1724	100%	0.0	99%	957/969
S. aureus RF122	1685	100%	0.0	98%	950/969
S. aureus ST398	1735	100%	0.0	99%	959/969
S. aureus M013	1701	100%	0.0	98%	953/969
S. aureus M1	1707	100%	0.0	98%	954/969
S. aureus Mu3	1696	100%	0.0	98%	952/969
S. aureus str. Newman	1707	100%	0.0	98%	954/969
S. aureus NCTC8325	1707	100%	0.0	98%	954/969
S. aureus VC40	1707	100%	0.0	98%	954/969

Table 43. *SAR0201* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0201	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	3280	100%	0.0	100%	1776/1776
S. aureus RN4282	3164	100%	0.0	99%	1755/1776
S. aureus T1	3280	100%	0.0	100%	1776/1776
S. aureus COL	3169	100%	0.0	99%	1756/1776
S. aureus MSSA476	3164	100%	0.0	99%	1755/1776
S. aureus MW2	3164	100%	0.0	99%	1755/1776
S. aureus N315	3169	100%	0.0	99%	1756/1776
S. aureus Mu50	3169	100%	0.0	99%	1756/1776
S. aureus T0131	3280	100%	0.0	100%	1776/1776
S. aureus TW20	3280	100%	0.0	100%	1776/1776
S. aureus BMB9393	3275	100%	0.0	99%	1775/1776
S. aureus LGA251	3208	100%	0.0	99%	1763/1776
S. aureus RF122	3197	100%	0.0	99%	1761/1776
S. aureus ST398	3192	100%	0.0	99%	1760/1776
S. aureus M013	3181	100%	0.0	99%	1758/1776
S. aureus M1	3169	100%	0.0	99%	1756/1776
S. aureus Mu3	3169	100%	0.0	99%	1756/1776
S. aureus str. Newman	3169	100%	0.0	99%	1756/1776
S. aureus NCTC8325	3169	100%	0.0	99%	1756/1776
S. aureus VC40	3169	100%	0.0	99%	1756/1776

Table 44. *SAR0174* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0174	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1801	100%	0.0	100%	975/975
S. aureus RN4282	1679	100%	0.0	98%	953/975
S. aureus T1	1801	100%	0.0	100%	975/975
S. aureus COL	1674	100%	0.0	98%	952/975
S. aureus MSSA476	1679	100%	0.0	98%	953/975
S. aureus MW2	1679	100%	0.0	98%	953/975
S. aureus N315	1659	100%	0.0	97%	950/975
S. aureus Mu50	1659	100%	0.0	97%	950/975
S. aureus T0131	1790	100%	0.0	99%	974/975
S. aureus TW20	1796	100%	0.0	99%	974/975
S. aureus BMB9393	1796	100%	0.0	99%	974/975
S. aureus LGA251	1652	100%	0.0	99%	948/975
S. aureus RF122	1668	100%	0.0	98%	981/975
S. aureus ST398	1768	100%	0.0	99%	969/975
S. aureus M013	1638	100%	0.0	97%	945/975
S. aureus M1	1674	100%	0.0	98%	952/975
S. aureus Mu3	1659	100%	0.0	97%	950/975
S. aureus str. Newman	1674	100%	0.0	98%	952/975
S. aureus NCTC8325	1674	100%	0.0	98%	952/975
S. aureus VC40	1674	100%	0.0	98%	952/975

Table 45. *SAR1495* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1495	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1647	100%	0.0	100%	906/906
S. aureus RN4282	1647	100%	0.0	100%	906/906
S. aureus T1	1647	100%	0.0	100%	906/906
S. aureus COL	846	74%	0.0	99%	524/557
S. aureus MSSA476	1452	98%	0.0	99%	855/891
S. aureus MW2	1447	98%	0.0	99%	855/891
S. aureus N315	1419	100%	0.0	99%	860/906
S. aureus Mu50	1419	100%	0.0	99%	860/906
S. aureus T0131	846	74%	0.0	99%	537/572
S. aureus TW20	486	74%	0.0	99%	524/557
S. aureus BMB9393	846	74%	0.0	99%	524/557
S. aureus LGA251	846	74%	0.0	99%	480/497
S. aureus RF122	863	61%	0.0	99%	527/557
S. aureus ST398	713	63%	0.0	99%	520/584
S. aureus M013	1447	98%	0.0	99%	855/891
S. aureus M1	846	74%	0.0	99%	524/557
S. aureus Mu3	1419	100%	0.0	99%	860/906
S. aureus str. Newman	486	74%	0.0	99%	524/557
S. aureus NCTC8325	846	74%	0.0	99%	524/557
S. aureus VC40	846	74%	0.0	99%	524/557

Table 46. *SAR0706* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR0706	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	195	100%	0.0	100%	105/105
S. aureus RN4282	195	100%	0.0	100%	105/105
S. aureus T1	195	100%	0.0	100%	105/105
S. aureus COL					
S. aureus MSSA476	191	100%	0.0	99%	101/105
S. aureus MW2					
S. aureus N315					
S. aureus Mu50					
S. aureus T0131					
S. aureus TW20					
S. aureus BMB9393					
S. aureus LGA251					
S. aureus RF122					
S. aureus ST398					
S. aureus M013					
S. aureus M1					
S. aureus Mu3					
S. aureus str. Newman					
S. aureus NCTC8325					
S. aureus VC40					

Table 47. *SAR1288* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1288	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	660	100%	0.0	100%	357/357
S. aureus RN4282	649	100%	0.0	100%	357/357
S. aureus T1	660	100%	0.0	100%	357/357
S. aureus COL	154	41%	0.0	92%	101/110
S. aureus MSSA476					
S. aureus MW2					
S. aureus N315					
S. aureus Mu50					
S. aureus T0131	649	100%	0.0	99%	355/357
S. aureus TW20	649	100%	0.0	99%	355/357
S. aureus BMB9393	649	100%	0.0	99%	355/357
S. aureus LGA251					
S. aureus RF122					
S. aureus ST398	649	100%	0.0	99%	355/357
S. aureus M013					
S. aureus M1	154	41%	0.0	992%	101/110
S. aureus Mu3					
S. aureus str. Newman					
S. aureus NCTC8325					
S. aureus VC40					

Table 48. *SAR1558* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1558	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	804	100%	0.0	100%	435/435
S. aureus RN4282	654	100%	0.0	94%	408/438
S. aureus T1	654	100%	0.0	94%	408/438
S. aureus COL					
S. aureus MSSA476	654	100%	0.0	94%	408/438
S. aureus MW2	793	100%	0.0	99%	433/435
S. aureus N315					
S. aureus Mu50					
S. aureus T0131					
S. aureus TW20					
S. aureus BMB9393					
S. aureus LGA251					
S. aureus RF122					
S. aureus ST398					
S. aureus M013					
S. aureus M1					
S. aureus Mu3					
S. aureus str. Newman					
S. aureus NCTC8325					
S. aureus VC40					

Table 49. *SAR2496* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR2496	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	2859	100%	0.0	100%	1548/1548
S. aureus RN4282	2772	100%	0.0	99%	1551/1548
S. aureus T1	2859	100%	0.0	100%	1548/1548
S. aureus COL	2772	100%	0.0	99%	1535/1551
S. aureus MSSA476	2772	100%	0.0	99%	1535/1551
S. aureus MW2	2787	100%	0.0	99%	1535/1548
S. aureus N315	2771	100%	0.0	99%	1532/1548
S. aureus Mu50	2771	100%	0.0	99%	1532/1548
S. aureus T0131	2772	100%	0.0	99%	1535/1551
S. aureus TW20	2772	100%	0.0	99%	1535/1551
S. aureus BMB9393	2772	100%	0.0	99%	1535/1551
S. aureus LGA251	2782	100%	0.0	99%	1534/1548
S. aureus RF122	2798	100%	0.0	99%	1537/1548
S. aureus ST398	2809	100%	0.0	99%	1539/1548
S. aureus M013	2782	100%	0.0	99%	1534/1548
S. aureus M1	2772	100%	0.0	99%	1535/1551
S. aureus Mu3	2771	100%	0.0	99%	1532/1548
S. aureus str. Newman	2743	100%	0.0	99%	1529/1550
S. aureus NCTC8325	2772	100%	0.0	99%	1535/1548
S. aureus VC40	2772	100%	0.0	99%	1535/1548

Table 50. *SAR1831* gene multiple sequence BLAST, pairwise alignment parameters and comparisons for *S. aureus* MRSA252 sequence with another 19 *S. aureus* strains.

SAR1831	**Max score**	**Query cover**	**E-value**	**Similarity (%)**	**Similarity (base pairs)**
S. aureus MRSA252	1563	100%	0.0	100%	846/846
S. aureus RN4282	1502	100%	0.0	99%	835/846
S. aureus T1	1563	100%	0.0	100%	846/846
S. aureus COL					
S. aureus MSSA476					
S. aureus MW2					
S. aureus N315	1502	100%	0.0	99%	835/846
S. aureus Mu50					
S. aureus T0131	1550	100%	0.0	99%	844/846
S. aureus TW20	1557	100%	0.0	99%	845/846
S. aureus BMB9393	1563	100%	0.0	100%	846/846
S. aureus LGA251					
S. aureus RF122					
S. aureus ST398					
S. aureus M013					
S. aureus M1					
S. aureus Mu3					
S. aureus str. Newman					
S. aureus NCTC8325					
S. aureus VC40					

Appendix 2.

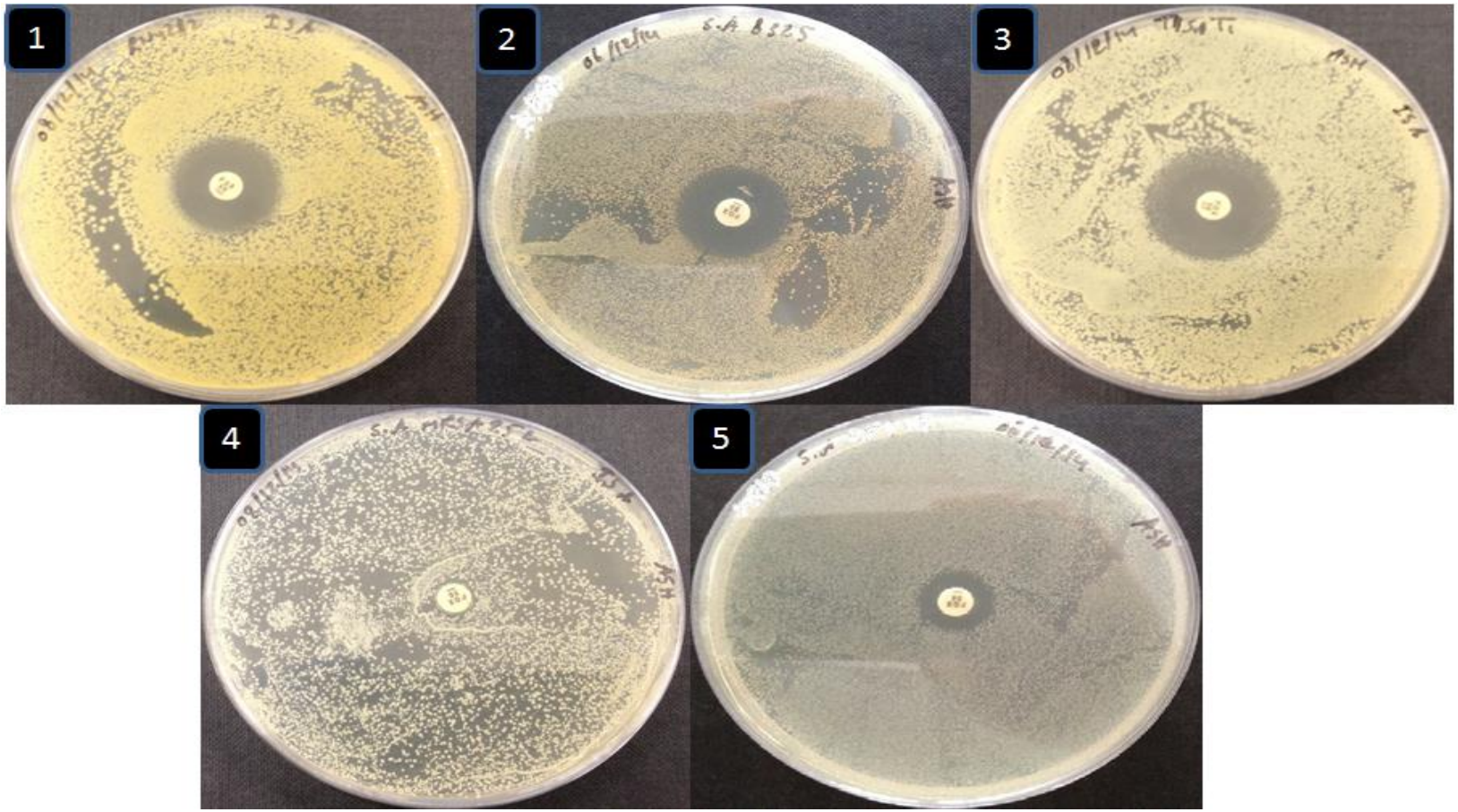

Figure 1. Cefoxitin disk diffusion plate test showing methicillin-resistance in *S. aureus* strains

Test was performed with 30 µg of cefoxitin per disk on 25 ml Iso-Sensitest agar incubation at 35°C for 18-20 h. Zone size was interpreted according to the CLSI as follows, susceptible, ≥ 22 mm; and resistant, ≤ 21 mm. 1. *S. aureus* RN4282: 21mm, 2. *S. aureus* ACTC8325: 21mm, 3. *S. aureus* T1: 25mm, 4. *S. aureus* MRSA252: 0 mm, 5. *S. aureus* NCTC12493 (control strain): 10mm.